Qohelet Euporia: a Domain-specific Language for the Encoding of the critical Apparatus

Luigi Bambaci
Dipartimento di
Civiltà e Forme del Sapere
Università di Pisa
Pisa, Italy
luigibambaci@yahoo.it

Federico Boschetti
CoPhiLab
Istituto di Linguistica Computazionale
"A. Zampolli", CNR
Pisa, Italy
federico.boschetti@ilc.cnr.it

Riccardo Del Gratta
LaRI
Istituto di Linguistica Computazionale
"A. Zampolli", CNR
Pisa, Italy
riccardo.delgratta@ilc.cnr.it

***Abstract*—Encoding multilingual variant readings is time-consuming and error-prone. The guidelines provided by the Text Encoding Initiative (TEI) ensure data interchange, but the XML-TEI verbosity is at risk of distracting annotators with a traditional background in philological studies from their critical activity. We illustrate how a Domain-specific Language (DSL) facilitates both the manual annotation of the critical apparatus and the data interchange. Our case study is based on the multilingual tradition of the biblical book of Qohelet, which has been annotated through the annotation tool based on DSLs named Euporia.**

***Index Terms*—digital philology, biblical studies, digital scholarly editing, textual scholarship, ecdotics, textual criticism, Old Testament studies, Hebrew Bible, XML-TEI textual encoding, computer-assisted textual criticism**

I. Introduction*

The activity of the textual philology is mainly a comparative activity, as stated by Cerquiglini. It compares manuscripts or printed editions of a given work (the *witnesses*), in order to detect differences (*readings* and *variants*).[1] In collating witnesses, the philologist moves further along large sections of the text without encountering differences. In these cases, the assumption is that the text of the work has been transmitted faithfully. When a variant arises, the philologist assumes an alteration of the textual structure. As pointed out by Segre, different alterations highlight a *diasystem*, a set of different textual systems, the one of the text and the ones of copyists-tradents.[2] The variant readings are the clues through which it is possible to infere such a diasystem, to study the textual history of the text (its *tradition*), and to try to reconstruct the earliest attainable form, removing errors due to the copying process and selecting the contextually more suitable readings which are likely to be original. The job of the philologist, therefore, lies in detecting the variants, in evaluating them and in making a choice: variants which have more chance to be original are placed inside the critical text, excluded variants are recorded in the critical apparatus.

The critical apparatus is the part, usually placed at the foot of the page, in which the editor gathers, mainly, readings taken from witnesses and conjectural emendations proposed by scholars.[3] Despite the hierarchical prominence of the critical text towards the critical apparatus, only this latter leaves traces of the reconstruction process, summarizing the diasystem of the tradition and carrying out a full assessment of the readings: as pointed out by Buzzoni, "[i]t is therefore in the apparatus that the diasystem of the tradition is best highlighted, and its *historicity* fully appreciated. [...] the critical apparatus is the key that allows the reader to understand the choices made by the editor to present the text in that particular shape. It is in the apparatus that the reader finds information about the editorial process that resulted in the text he or she is reading — thus enabling her/him to evaluate the editor's decisions — as well as the different shapes assumed by the text itself in the period in which it was composed and committed to posterity."[4]

The logic underlying the preparation of the critical apparatus

*Section I, II, III, IV were written by Luigi Bambaci; section V by Federico Boschetti and section VI by Riccardo Del Gratta.

[1] Cerquiglini [1] 37: "Philology, created for editing the ancient and sacred Latin and Greek works that were reproduced especially during the Middle Ages, is a measured and patient practice of comparison; it compares manuscripts separated only — this is axiomatic — by the changes specific to the act of copying. When *tradition* (i.e., all of the manuscripts that have come down to us) presents different readings (i.e., lessons from *lectio*: what one reads) at a certain point in the text, there is a *variant* (philology sometimes calls it *innovation*, as a reproach), and one needs to make sure which is the good text (for the "good reading" etc.)." For a definition of the terms "reading" and "variant" cf. Gerd [2] 28: "A reading is the generic term for the wording of a textual unit in which a manuscript is distinguished from one or more or from all other manuscripts. A *variant* refers to one of at least two readings of the same textual unit which is grammatically correct and logically possible." Cf. also Epp [3] 57 ff. In Old Testament text-critical studies, the term reading defines all details in manuscripts, while readings differing from the Hebrew text (the so-called Masoretic Text) are named variants, cf. Tov [4] 430: "All elements in the text are named readings, and similarly all details in Mss that differ from a given yardstick are called variant readings, that is, readings which are at variance with the base text. In the case of the O[ld] T[estament], M[asoretic] T[ext] is taken as the base for all comparisons, so that all details which differ from MT are called variant readings." Similarly Tov [5] 266.

[2] Segre [6] 14, 58-9.

[3] Cf. Fränkel [7] 10-16, Avalle [8] 122-4.

[4] Buzzoni [9] 64.

is a matter of editorial choices. The editor can decide whether to prepare the apparatus of a collation, and hence to record the whole amount of textual variants, or the apparatus of a critical edition, which consists of a selection of the most significant instances.[5] The editor can choose to record only substantial variants (variants which are considered to affect significantly the meaning of the work) and to leave out formal variants and accidentals, such as those concerning orthography and punctuation.[6] Once the critical apparatus has been prepared, the scholar can decide whether to analyse the gathered data according to specific needs: significant variants can be selected for establishing the *stemma codicum*, a graphical depiction in form of genealogical tree which represents the hierarchical relations between witnesses;[7] statistical-based analysis can also be performed, such as clustering or cladistics;[8] variants of single witnesses can be gathered and studied independently, in order to assess the textual value of their readings separately; variants corresponding to certain categories can be collected, in order to study their frequency within the textual tradition and to prepare repertories of copy errors; similarly, conjectural emendations can be extracted and repertories can be be prepared.

The language of critical apparatuses conveys information by means of two main tools: abbreviations and symbols (including numbers and punctuation) and the position of textual elements. The firsts may concern witnesses (which are recorded with conventional *sigla*), evaluation of variants (*fac* for "facilitation", *assim* for "assimilation"), features about the representation of sources (*sup ras* for "erasure above", *primo* for "first copyist's hand"), the location in the text (expressed by number of chapter, verse, paragraph), further editorial interventions (such as asterisks for corrupted passages, angle brackets for integrations) and so on.[9] The position regards the status of textual elements: thus, for example, the first word of the apparatus entry (eventually separated by a square bracket or a double point) is the word of the critical text for which a variant or a conjecture is given, while the strings after it are the variants or the conjectures; a list of witnesses *sigla* may mean that they share the same reading, the same phenomenon of textual variation, and so on. The structure of a critical apparatus is meant to be an economic solution to the verbosity of the natural language.[10] The language of the critical apparatus, therefore, can be considered as an *artificial* (or *planned*) language.[11] Inasmuch it exploits symbols and a conventional vocabulary, it is comparable to the languages of mathematics or chemistry, intended as "nonredundant, formulaic or symbolyc languages to facilitate scientific thought."[12]

The information structured in this way is implicit: where the user accustomed to philological conventions reads, by way of inferences, a set of meaningful and coherent philological data, the computer "reads" a succession of strings, integers and white spaces. In order to enable the computer to process such information, it must be explicit. A way for render it explicit is to mark-up the text, that is, to apply a set of markers (or tags) which describe the editor's interpretation of textual phenomena. The process of inserting such explicit markers for implicit textual features, named *textual encoding*, can be performed by using the so-called mark-up languages, such as XML.

One of the main activities of digital philologists is the encoding of variant readings and conjectures, in order to record the differences among the witnesses or the emendations suggested by the scholars. The encoded variants should not be just machine readable, as they are in a digitized critical apparatus acquired from a printed edition and rendered on the screen in the same way of the original paper version. On the contrary, they should be fully machine actionable, in order to allow the creation of dynamic apparatus by the application of filters, the visualization through complex graphs and the construction of textual indexes and concordances based not only on the reference editions but also on their variants. The guidelines provided by the *Text Encoding Initiative* (TEI) related to the critical apparatus are the result of a collective effort within the community of digital humanists. They provide a mark-up vocabulary for a variety of problems arising in textual criticism, with a large coverage of different usage cases. TEI pursues the standardization of markup schemas and vocabulary for literary and philological studies, thus ensuring data interchange. More than other XML vocabularies, TEI markup schemas meet scholars' need to encode texts that can be reused as a starting point for further inquiries.

The TEI guidelines are flexible enough to provide the user with three different annotation strategies, in order to link the critical apparatus to the text: the location-referenced method, the double-end-point-attached method and the parallel segmentation method.[13] The first method offers a solution suitable for the encoding of printed critical apparatuses. Being linked to

[5]Cf. Boschetti [10]: "[...] The critical apparatus is a selection. If the text accepted by the editor is subjective in its substitutions, the critical apparatus is subjective in its omissions. The critical apparatus [...] can be considered as an anthology, not as an exhaustive repertory of information. Only collations and repertories of conjectures can claim completeness, even if the former is limited to the number of examined manuscripts and the latter to the number of examined printed editions, commentaries and articles."

[6]On the distinction between substantives and accidentals cf. Contini [11] 38 ff. According to Contini, Gaston Paris was the first to distinguish between *critique des formes* and *critique des leçons* in his introduction to *Vie de saint Alexis* (1872), cf. Reeve [12] 61 ff. Such distinction is also found in the work of Greg [13], theoretician of the so-called copy-text method, cf. Greetham [14] 333 ff.

[7]Maas [15] pg. 14 § 21: "The diagram which exhibits the inter-relationship of the witnesses is called *stemma*. The image is taken from genealogy: the witnesses are related to the original as the descendants of a man are related to their ancestor." Cf. also Avalle [8] 97-98.

[8]For a summary of the statistical techniques applied to textual traditions cf. Hockey [16] 144 f. and Pierce [17]. On cladistics analysis see the two volumes of *Studies in Stemmatology*, Reenen et al. [18] and [19].

[9]Maas [15] 15 ff., Avalle [8] 123-4.

[10]Pasquali [20] 52: "the use of symbols [...] is intended to quickly indicate what we are talking about, without having to start the explanation all over again; it aims, therefore, [...] to a purely economic purpose."

[11]Cf. Blanke [21], Libert [22].

[12]Blanke [23] 33.

[13]TEI Consortium, eds. "12.2 Linking the Apparatus to the Text." *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. [3.5.0.]. [29th January 2019]. TEI Consortium. http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DS.html#DSFLT ([10/03/2019]).

the base text by means of annotations specifying the point on which variation insists, it can be stored separately from it (*external* apparatus). The main drawbacks of such a method are represented by the lack of precision in indicating the exact word-token interested by textual variation. Unlike the location-referenced method, the linkage system of the double-end-point-attached method is mainly based on milestones elements. This allows not only a far more precise identification of the variant units, but also enables to handle the problem of the overlapping variants. This extreme flexibility and precision, however, is counterbalanced by the objective difficulty of performing a manual encoding and of reading and interpreting the encoded file without mechanical assistance. The parallel segmentation method is based on an *in-line* approach and does not necessarily depend on the concept of base text. It is therefore optimal when we lack a reference edition. In this method, each segment of the critical text and corresponding variants are synchronized with one another. This permits the comparison of many spans from different witnesses and is therefore suitable when one wishes to present parallel texts. It is more precise in selecting the variant units than the first, and far easier to be implemented by hand than the second. These features make it the preferred one among the community of digital scholars and the most widely-adopted in many digital-born, TEI compliant projects. The main drawback is represented by the overlaps of variants. In order to avoid it, the editor is compelled to conflate all the overlapping variants in a single reading[14] into pieces. Such a fragmentation of the logic order of variants, in many cases, does not fit well with the way innovations in copying and transmitting texts are normally performed, and may thus lead to an ambiguous and inappropriate representation of the textual variation in the critical apparatus.

The guidelines also describe how to widen the TEI schema itself using new tags and attributes[15] or, on the contrary, how to narrow it by the definition of restrictive schemas,[16] in order to limit the ambiguities and improve the interoperability.[17] The compliance to the TEI guidelines is among the best practices in digital philology. Indeed, academic courses and workshops regarding how to annotate digital scholarly editions through the XML-TEI mark-up language are more and more frequent. The great advantages of manual encoding, such as the portability (the independence from hardware and software components[18]), constitute a severe limitation for the user with no technological background. Moreover, the manual annotation of variant readings is a time-consuming, error-prone, and non-trivial task, especially when dealing with very rich and complex textual traditions. As we will discuss in section II, there are several methods of facing the problem of a manual encoding: the implementation of graphical interfaces, the use of abbreviated markers and, finally, the annotation through a Domain-specific Language, a programming language designed for implementing domain-related specific tasks (cf. section III). In section IV we expose the results, based on the ongoing multilingual critical apparatus of Qohelet, and we discuss them in section V. In section VI, finally, we sketch up what we are planning as next steps.

II. Background

Tools both for encoding and annotating literary texts and either for the visualization or publication of digital scholarly editions are currently available. Among the latter, the Critical Apparatus Toolbox (CAT [27]) and Edition Visualization Technology (EVT [28]). Both the applications allow the user to visualize (CAT) or publish (EVT), through the parallel segmentation method, texts encoded in XML-TEI.

The common trend of the available encoding tools moves towards a simplification of the XML manual annotation, by means of user-centered graphic interfaces or thanks to a simplification of the tagging process. Among the integrated development environments intended mainly for textual-critic activity, it is worth mentioning the Cooperative Web-Based Editor for Critical Editions (CEED [29]), which provides a user-friendly graphic interface for the encoding of the variant readings. The application, therefore, is conceived also for users with little or no knowledge of the technical aspects of the TEI encoding. By making the mark-up process automatic, the graphic interface ensures the possibility of avoiding mark-up syntactic errors. Nevertheless, since the aim is to cover the richness of the mark-up potentialities offered by TEI, this user interface could turn out to be difficult to handle for the philologist, and might give the impression of a lack of control over the text to be encoded.

Conceived mainly for the editing of papyrological texts is the Papyrological Editor (PE [30]), available on Papyri.info.[19] The encoding process is facilitated by a plain graphic interface as well as by an annotating system that simplifies, by means of abbreviations, the form of TEI markers. Such an encoding system, by combining both the expressivity of the XML language and the usual conventions of the textual criticism, is closer to the practices of the domain specialists.

Finally, tools for semantic annotation, such as Pundit [31] [32], allow the integration with Linked Open Data, according to the paradigm of the semantic web.

Our DSL-based approach is intended to be an alternative to both the manual encoding and to the encoding carried out through GUI or by means of shortened tags. As stated in section I, the language of critical apparatuses is a designed, artificial language. The methods editors employ when

[14] Epp [3] 60: "A "variation-unit" is that determinate quantity or segment of text, constituting a normal and proper grammatical combination, where our MSS [manuscripts] present at least two variant". An alternative term is "variant location", cf. De Vos et al. [24] 113: "A variant location is a locus in the text where at least two concurrent readings exist". The terms *variation place*, *variant place* or *place of variation* are also used, cf. Salemans [25] 23.

[15] For the TEI customization, see ⟨http://www.tei-c.org/release/doc/tei-p5-doc/en/html/USE.html⟩.

[16] For instance, see the TEI subset of EpiDoc: ⟨https://sourceforge.net/projects/epidoc⟩.

[17] About the difference between interchange and interoperability related to the TEI guidelines, see ⟨https://bit.ly/2vYd0zw⟩.

[18] Cf. Ciotti [26] 24-25.

[19] ⟨http://papyri.info⟩.

recording variants in critical apparatuses vary from edition to edition. It may depend on the features of the textual tradition under consideration, on the different theoretical conceptions about the genesis and evolution of the literary texts, on scholarly orientations and current trends. From a historical point of view, as pointed out by Kenney, the birth of the modern critical apparatus is to be placed at the end of the 18th cent., within the field of New Testament textual criticism, with the editions of Bengel (1734) — who can be given credit of having coined, probably, the term *apparatus criticus*[20] — and Wettstein (1751-2).[21] It was the work of these textual critics that gave rise to the modern conception of critical apparatus, intended as a concise system of annotation, separated from the textual commentary, of manuscript *sigla* and corresponding readings. According to Kenney, this tendency to abstraction and to the employment of conventional symbols and abbreviations was not the norm in philological studies[22] and was opposed until recent times.[23] A language of this sort, in which all the constituents are defined in a concise, nonredundant and unambiguous way, is a *formalized* language. To formalize a language is a matter of constructing its syntax and indicating its semantics.[24] The formalization implies that each apparatus component is assigned to a specific type with a specific meaning and that rules of formation of valid expressions are established. As we will see in the following section, it is possible to prepare a critical apparatus which reflects these features. The main purpose is to allow the computer to interpret it and to interact with it (cf. sections III and V). In this respect, the formalized language of the critical apparatus will function as a sort of programming language. Unlike a general-purpose programming language, it is domain-specific: a language of limited expressiveness optimized for a particular domain of knowledge or domain of application,[25] which is, in our case, the ecdotic. In order to allow such an interaction, we wrote the formal grammar. The grammar lies down the rules in order to allow the parser created with ANTLR software[26] to analyse and recognize automatically the structure of the critical apparatus and all its elements. The syntactic tree generated by the parser is traversed by the listener, which translates the input model created by the parser to an output with print statements, that is, TEI corresponding sequences of markers and attributes. It is, therefore, a substitutive, model-driven translational process:[27] from a given input (the apparatus components interpreted by the parser) to the desired output (appropriate XML-TEI tags).

Our case study concerns the book of Qohelet, one of the book of the Hebrew Bible. Euporia Qohelet is a project which aims to produce a native digital eclectic edition of that the book.[28] At present, the two more recent scholarly editions (Rudolph et al. [41], Schenker et al. [42]) follow the diplomatic model. Neither a complete collation of the witnesses of that book nor a native digital scholarly edition of the Hebrew Bible has been published. Extant electronic versions available on the Internet or in commercial computer programs – *Accordance*,[29] *Bibleworks*,[30] *Logos*[31] – are indeed not critical: they simply provide the text of one codex (mainly the most used base text, the *Codex Leningradensis*) and a limited set of ancient versions (mainly Greek, Latin and Targumim), "and therefore have no added value relating to their Editionstechnik."[32] Also the electronic versions (when available) of the aforementioned editions "do not reflect a decision making process, since they simply continue the production line of existing paper editions." Projects that aim at producing a multi-column representation of the witnesses exist only on paper — so the *Synoptic Electronic Database* (SED) and *The Madrid Project of the Historical Books* — or are limited to few sources – so the *Computer Assisted Tools for Septuagint Studies* (CATSS) available on *Accordance* (Hebrew and Greek text only).[33] A project of a digital multiple-version edition, as attested by the numerous calls of scholars for such a project,[34] is therefore a *desideratum* and may represent a sound compromise for both who support the diplomatic method and who support the eclectic method: as pointed out by Tov "a combined diplomatic and eclectic edition will educate the users towards an egalitarian approach to the textual witnesses, combining the best of both systems."[35] As stated by Hendel, the digital medium will make possible "a wider distribution of knowledge and, one may hope, new kinds of textual scholarship. At a time when the humanities are in decline in its long trajectory since the Renaissance, the powers of philology may yet surprise us. With a new medium, whose entailments and implications are still being explored, we may be able to reimagine the axis of innumerable relationships in a very old book."[36]

III. Method

Euporia,[37] the annotation tool based on DSLs developed at the CoPhiLab of the CNR-ILC, has been formerly used

[20] Cf. Timpanaro [33] 65 n. 16 and Kenney [34] 294 n. 22. See also the historical discussion in Gane [35].

[21] Cf. Meztger [36] 48, 158 ff., Aland et al. [36] 8 ff., 72 ff.

[22] In the same Lachmann's edition of Lucretius' *De rerum natura* (1850) the critical notes were not gathered in a critical apparatus, but mixed within the exegetical notes of the *Commentarius*, cf. Gane [35] 23.

[23] The traditional conception of the classic studies as *bonae literae*, according to Kenney [34] 205, inspired distrust of all that is shortened, technical and algebric: "This kind of attitude still persists" — Kenney writes — "nowadays, his equivalent can perhaps be seen in the reluctance of some philologists to deal with techniques transferred from natural sciences and mathematics to literary studies."

[24] Cf. Grishin [37] 61.

[25] Fowler [38] 28.

[26] Parr [39].

[27] Cf. Parr [40] 295 ff.

[28] At present, the first three chapters have been encoded.

[29] https://www.accordancebible.com/.

[30] https://www.bibleworks.com/.

[31] https://www.logos.com/.

[32] Tov [43] 87.

[33] Cf. Tov [43] [44].

[34] Tigchelaar 2002 [45], Tov 2008 [46], Hendel 2008 [47], Segal 2017 [48]

[35] Tov [5] 365.

[36] Hendel [49]31-2.

[37] ⟨http://www.himeros.eu/euporia⟩.

for interpretative tasks, such as the identification of ritual frames in the ancient Greek tragedies documented in Mugelli et al. [50]. The work-flow of the study can be summarized as follows. The creation of the critical apparatus and the reconstruction of the critical text was preceded by a preliminary stage of analysis of the traditional critical apparatuses in the domain of Old Testament studies and the investigation of the best practices to render such information in XML-TEI. Among the available critical editions of the book of Qohelet, the critical apparatus of the Biblia Hebraica Quinta (BHQ, [42]), shaped on the one devised by CTAT Committee,[38] was selected as the best solution for the encoding of the new digital apparatus, for three main reasons. First, the BHQ represents the most recent edition of the book. Second, unlike the apparatuses of the other editions which are centred mainly on the variant readings diverging from the base text — the text of the *Bombergiana* in the *Biblia Hebraica* (BH, [52]) and the *Leningradensis* in the *Biblia Hebraica Stuttgartensia* (BHS, [41]) — the critical apparatus of the BHQ is positive: it records both the deviations from the base text and the readings supporting it, and it is, therefore, more complete in terms of information. Third, the structure and morphology of the BHQ critical apparatus is very rigorous and recursive, and, therefore, more suitable for an automatic analysis. An example of BHQ's critical apparatus is shown in Fig. 1. Here, after the number

8 וְזָהָב G^{Mss} S T (assim-ctx?) | καί γε ἀργύριον καὶ χρυσίον G* | V (indet)

Figure 1. Qohelet 2:8 (*Biblia Hebraica Quinta*)

indicating the verse of the chapter, there is the Hebrew word of the base text (the *Codex Leningradensis*) for which variants are attested (in TEI terminology, the *lemma*).[39] The reading of the lemma is supported by several witnesses of the Greek tradition (*siglum* "G^{Mss}"), the Syriac Version ("S") and the Aramaic Version (the Targum, "T"). In brackets, the editor expresses his evaluation on the readings of these witnesses ("assim-ctx", that is, assimilation to the context), which is uncertain (marked with "?"). After the lemma group, there are other two groups, each separated by a vertical line: the one attesting the variant of other Greek witnesses ("G") considered by the editor as representing the original Greek reading ("*") and, finally, the reading of the Latin Version, (the Vulgate, "V"), which the editor judged indeterminate ("indet"), that is, impossible to evaluate.

A possible conversion of this apparatus in an XML-TEI format, according to parallel segmentation method, is shown in Fig. 2. As it can be seen, a semantic function has been attached to each relevant element through a set of markers. In this case, the element `<app>` indicates the beginning of the apparatus entry containing both the lemma found in

```
1  <app>
2    <lem wit="#L #GMss #S #T"
3         cause="assim-ctx"
4         cert="unknown"><w>וזהב</w></lem>
5    <rdg wit="#G*"><w>καί</w>
6                   <w>γε</w>
7                   <w>χρυσίον</w></rdg>
8    <rdg wit="#V" ana="#indet"/>
9  </app>
```

Figure 2. Qohelet 2:8 (TEI compliant critical apparatus)

the reference text (`<lem>`) and the variants (`<rdg>`). Other information is encoded through the attributes. Thus, for example, the attribute `@wit` in `<lem>` contains the *sigla* of the witnesses (`#L #GMss #S #T`),[40] the attribute `@ana` (standing for analysis)[41] and `@cert` (certainty)[42] contain, respectively, the critical evaluation (`assim-ctx`, `#indet`) and the degree of likelihood (`unknown`).

Besides recording and evaluating variant readings, another important task of the textual critic is to choose those readings which, according to the editor's judgment, are likely to be original. An example of such philological procedure is shown in the critical apparatus of Fig. 3. The editor of the BHQ, after

הוֹלֵלוֹת (interp) | παραβολὰς G (S) (interp) | πλάνας α' V (interp) | παραφοράς θ' | והולחולתא דמלכותא T* | והולהולתא דמלכותא T (err) || pref וְהוֹלֵלוֹת (origin)

Figure 3. Qohelet 1:17 (*Biblia Hebraica Quinta*)

having presented the versional evidence, proposes at the end of the apparatus entry to choose a reading making it preceded by the abbreviation "pref" (short for "preferred readings"[43]). A way of encoding it is shown in Fig. 4. The element `<rdgGrp>`, which allows to group the variants for whatever theoretical

[38]Cf. Barthélemy [51] cf. Schenker et al. [42] XII.

[39]Cf. TEI Consortium, eds. "12.1.2 Readings." *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. [3.5.0.]. [29th January 2019]. TEI Consortium. http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DS.html#DSFLT ([10/03/2019])

[40]The asterisk expresses a reference to a list of witnesses (`<listWit>`) encoded previously in the XML-TEI file and providing all the relevant information about sources, cf. TEI Consortium, eds. "12.1.4.3 The Witness List." *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. [3.5.0.]. [29th January 2019]. TEI Consortium. http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DS.html#DSFLT ([10/03/2019])

[41]Cf. TEI Consortium, eds. "17.2 Global Attributes for Simple Analyses." *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. [3.5.0.]. [29th January 2019]. TEI Consortium. http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DS.html#DSFLT ([10/03/2019])

[42]Cf. TEI Consortium, eds. "21.1.2 Structured Indications of Uncertainty." *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. [3.5.0.]. [29th January 2019]. TEI Consortium. http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DS.html#DSFLT ([10/03/2019]).

[43]Schenker et al. [42] 17. The role played by the preferred readings in the BHQ and, more in general, in the diplomatic editions of the Hebrew Bible, is ambiguous: on one side, they are considered by the editor as superior to the transmitted text; on the other, they remain confined to the critical apparatus, where they appear mixed together with secondary readings. This ambiguity, rightly criticized by many scholars (cf. e. g. Borbone [53], Hendel [47], Tov [5] 360), is resolved in an eclectic edition, which places the preferred readings in a critically reconstructed text.

```
<app>
  <lem cause="interp"><w>הֹוֵללוֹת</w></lem>
  <rdgGrp cause="interp">
    <rdg wit="#G"><w>παραβολὰς</w></rdg>
    <rdg wit="#S"/>
  </rdgGrp>
  <rdg wit="#GAq #V" cause="interp"><w>πλάνας</w></rdg>
  <rdg wit="#GTh" cause="interp"><w>παραφοράς</w></rdg>
  <rdg wit="#T*"><w>וחולחולתא</w> <w>דמלכותא</w></rdg>
  <rdg wit="#T" cause="err"><w>והולהולתא</w> <w>דמלכותא</w></rdg>
  <rdg type="pref"><w>וְהֹוֵללוֹת</w></rdg>
</app>
```

Figure 4. Qohelet 1:17 (TEI compliant critical apparatus)

reason,[44] was used here to express the editor's evaluation of variants and to group readings according to shared innovations towards the reference text.[45] The reading of witness S, indeed, is similar to that one of G, but with slight and negligible modification (this is the meaning of the brackets surrounding S[46]). The preferred reading has been encoded with a `<rdg>` element marked with the attribute `type="pref"`.

This and other similar instances have been analysed, in order to cover the greatest possible number of text-critical problems. Relying upon the model of the critical apparatus proposed by the editors of the BHQ, we defined a DSL which allows the user to encode variants in a language that is very close to the semi-structured language familiar to traditional philologists in preparing a printed apparatus, but that is, at the same time, as expressive and unambiguous as a digital apparatus encoded in XML-TEI.

The second step was writing the critical apparatus as plain text. An example of apparatus entry is shown in Fig. 5. Unlike BHQ, here the variants are fully recorded. The degree

3:16 הָרָשַׁע L] *iniquitatem* V | *iniquitas* Hy || subst sem n גברא חייבא T | ὁ ἀσεβής Compl., Ald., Gra, Ra. | ܚܛܝܐ P | [illegible] B1, C || subst sem n ὁ εὐσεβὴς A, B, S, Ti. | [illegible] Syr | ጸድቅ Ae | բարեպաշտն An || prp הָרָשָׁע Ehr(1914) Gol(2004) | הָרָשָׁע Gra(1871) Dri(1905) Bar(1908) Zap(1911) Wil(1922) Gal(1940)

Figure 5. Qohelet 3:16

of collation,[47] moreover, is far higher, since it also includes secondary translations from the Greek version (Armenian and Ethiopic). As with the critical apparatus of BHQ, our critical apparatus is positive. The main difference lies in the nature of the lemma: in the BHQ, which is a diplomatic edition, the lemma is always represented by a reading of the *Codex Leningradensis*, while in Euporia's Qohelet, which aims at publishing an eclectic edition with a critical text, the lemma is constituted by the text of those readings which have been judged as superior by the editor. All the variants supporting the preferred readings (the lemmas) are positioned on the left of the apparatus entry. Each reading is characterized by the *siglum* of the witness. Readings which are to be ruled out and which share the same features are grouped together. Each group is separated by a double vertical line and is introduced by an annotation which indicates the typology of variation (in this case, a substitution of noun with semantic change of meaning, "subst sem n").[48] As in the BHQ, at the end of the apparatus entry the readings which have been preferred by scholars are presented, taken either from other witnesses or reconstructed by conjecture. After each reading, bibliographical references (the name of the editor and the date of the edition, comment or article) are provided.

Such a critical apparatus has been written on Euporia's interface (Fig. 6). The choice of adopting the plain text was dictated by two factors, one theoretical and one practical. In the first instance, it allows the critical apparatus to be written without having to depend on a particular development environment and to be downloaded in it at a later stage as well. Secondarily, the independence from any specific input format, such as XML, allows the philologists to stay focused on given research tasks, writing the critical apparatus as they would have normally done in their customary research practice. For the same reason, we opted for retaining the long-established structure of printed critical editions, where two main textual flows can be distinguished: the critical text or the base text on one side and the critical apparatus on the other. In this way, the annotator can easily link his or her annotations to the reference text (in this case, the text of the *Codex Leningradensis* on the left of Euporia's interface), without having to handle long in-line annotated texts, which are usual for digital philologists accustomed to TEI encoding, but unfamiliar to traditional scholars.

The preparation of the critical apparatus is based on the Context-free Grammar (CFG), that defines the DSL. A CFG is a type of formal grammar which consists of a set of rules describing a formal language. The rules of the grammar enable the computer to parse it and to verify its correctness. Grammars, therefore, are real executable "programs" written in a DSL specifically designed for expressing language structures.[49] There are two kind of rules: rules for tokenization (token rules), which determine the vocabulary symbols (readings, *sigla* and so on), and rules for syntactic structure (parser rules), which determine the syntax (the position). Let us take the first part of the apparatus entry of the example shown in Fig. 5. It consists of four parts: the variation unit ("3:16", the location in the text body, expressed by the number of chapter and verse), the words constituting the lemma; the *siglum* of the witness ("L") and finally the square bracket that closes the lemma. The first operation to do is to tokenize, that is, to let the computer isolate each element from the

[44] [54] 469. Cf. cf. TEI Consortium, eds. "12.1.3 Indicating Subvariation in Apparatus Entries." *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. [3.5.0.]. [29th January 2019]. TEI Consortium. http://www.tei-c.org/release/doc/tei-p5-doc/en/html/DS.html#DSFLT ([10/03/2019])

[45] Cf. [42] XVI f., LXXIII.

[46] Schenker et al. [42] LXXIX-LXXVIII.

[47] The term is borrowed from Greg [55] 17, where it means " the minuteness of the variants of which notice is taken".

[48] The presentation of the typology of variants is shaped on that one provided by Catastini [56] 12 and [57] 37.

[49] Cf. Parr [40] 38.

Figure 6. Euporia Interface

textual flow. This operation is named *tokenization* or *lexical analysis*. The program that tokenizes is called *lexer*. Fig. 7 shows the rules of the CFG fit for purpose. The goal of the

```
1   grammar QoheletEuporia ;
2
3   app: NUM+ DOUBLE_POINT NUM+
4        HEBW+ ALPHA_SEQ R_BRACKET ;
5
6   NUM : [0-9]+('.'[0-9]+)? ;
7   DOUBLE_POINT : ':' ;
8   HEBW : [\u0590-\u05ff]+ ;
9   ALPHA_SEQ : [a-zA-Z]+ ;
10  R_BRACKET : ']' ;
```

Figure 7. Rules for tokenization

lexer is to emit a sequence of tokens. Each token has two primary attributes: a token type or class (symbol category) and the text associated with it: for instance, `NUM` allows to tokenize integers, `DOUBLE_POINT` and `R_BRACKET` the punctuation, `HBW` sets the Unicode characters of the Hebrew alfabet and `ALPHA_SEQ` the characters of the Latin alphabet. Once the lexer has processed characters, it passes tokens to the parser, which checks syntax and creates a parse tree. A parse tree, or syntax tree, shows how the parser recognized the structure of the input sentence with all its components. The result of applying the rules for tokenization is shown in the parse tree of Fig. 8. The `app` rule is the root node. The leaves of the parse tree are the input tokens. As it can be seen, the rule `app` defines the syntactic structure. An apparatus entry, indeed, always consists of: a sequence of numbers, which can be repeatable (as expressed by the subrule operator

Figure 8. Parse tree

"+"), a double point, another sequence of numbers, Hebrew characters, Latin characters and finally the square bracket.
The CFG enables also to attach labels to tokens, in order to remind their semantic function. In this case, the first number represents the chapter, while the second the verse; the double point and the square bracket are but separators; the alphabetic characters represent the words of the lemma (in Hebrew) or the *sigla* the witnesses (in this case, in Latin alphabet). Such information can be applied to the aforementioned rules through labels, as shown in Fig. 9. Here, all the rules beginning with lower cases are labels: the rule `w` (short for "word") labels the rule `HEBW`, thus defining every Hebrew token; to the double point the function of separator has been assigned through the rule `locSep`; chapter and verse are defined on the basis of their position: the first number always represents the chapter, the second always the verse, and both are always placed at the first position in the apparatus. The rule `lemma` allows to identify the location (chapter and verse) separately. The square bracket always indicates the end of the lemma (`lemSep`). Once they have been settled, these labels can be used elsewhere in the grammar as shorthands. The resulting tree is shown in Fig. 10. Once the apparatus and the lemma

```
grammar QoheletEuporia;

app : loc lem;
lem : w+ wit lemSep;
loc : chap + locSep + v?;
chap : NUM;
v : NUM;
locSep : DOUBLE_POINT;
lemSep : R_BRACKET;
wit : ALPHA_SEQ;
w : HEBW ;

NUM : [0-9]+('.'[0-9]+)?;
ALPHA_SEQ : [a-zA-Z]+;
DOUBLE_POINT : ':';
R_BRACKET : ']';
HEBW : [\u0590-\u05ff]+;
```

Figure 9. Syntax analysis

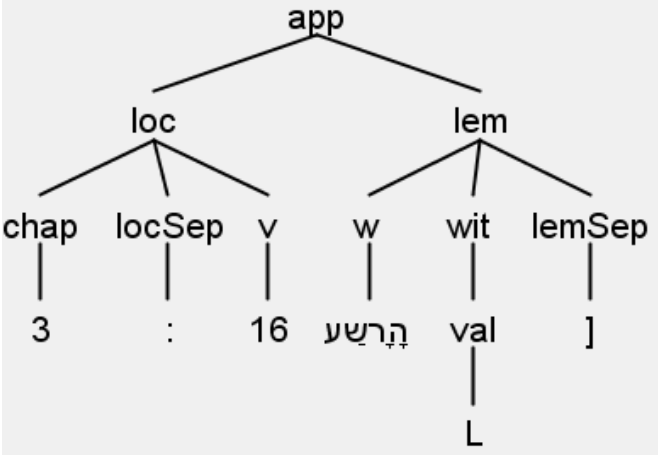

Figure 10. Parsing location and lemma

have been defined, CFG rules have been set in order to deal with readings and reading groups. CFG's potentialities and flexibility have proved to be optimal for such an undertaking. Thanks, indeed, to the recursive structure of grammatical rules, it is possible to express, through concise and simple definitions, all the necessary instructions for automatically parsing long list of variants. The rules are listed in Fig. 11. Inside `WIT_HEBR` and `WIT_GREEK` are recorded the *sigla* of the witnesses under consideration. This use is near to XML-TEI lists of witnesses (`<listWit>`). The rule `GRCW` defines the Greek Unicode characters. At the top of the grammar, the reading groups are described. The first rule means that a reading group can be composed by a separator (a double line), an element `ana` (short for "analysis" which was used for describing the typology of variation) and finally by one or more readings (`rdg`); a `rdg`, in turn, consists of a sequence of words (`w+`), of witnesses *sigla* and of a separator (a

```
rdgGrp : rdgGrpSep? ana? (rdg)+ ;
rdg : (w+)?  wit + rdgSep? ;
ana : ALPHA_SEQ+ anaSep ;
wit : (val)+ ;
w : (HEBW|GRCW) ;
val :  (WIT_HEBR|WIT_GREEK|WIT_TARG) ;

rdgSep : RDG_SEP ;
rdgGrpSep : RDGGROUP_SEP ;
anaSep : DOUBLE_POINT ;

ALPHA_SEQ : [a-zA-Z]+ ;
HEBW : [\u0590-\u05ff]+ ;
GRCW : [\u0370-\u03ff\u1f00-\u1fff]+ ;

WIT_HEBR : 'L' | 'Qa' | 'Qb' ;
WIT_TARG :'T' | 'Ts' | 'Tl10' | 'Urbl' ;
WIT_GREEK :'G'| 'Ra.' | 'Ti.' | 'Gra.' |
          'Compl.' | 'Ald.' | 'A'| 'B'| 'S' ;

DOUBLE_POINT: ':' ;
RDG_SEP: '|' ;
RDGGROUP_SEP: '||' ;
```

Figure 11. Rules for defining reading groups

single line); the words of the reading may be in Hebrew or Greek alphabet, and so forth, up to the end of the grammar, were token rules are placed. The result is shown in Fig. 12. The last elements to be defined are the preferred readings.

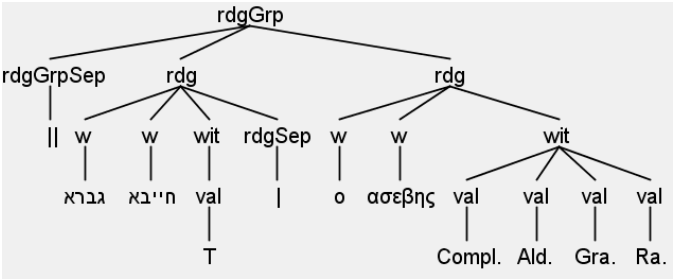

Figure 12. Parsing reading groups

The grammatical rules which define them are achievable by modifying the structure of rules described above, as shown in Fig. 13. Adding the "or" operator ("|") to the rule `rdgGrp` and `rdg`, indeed, it is possible to customize new typologies of reading and reading groups. In this case, it was specified that a reading group may be introduced whether by the analysis or by the `type` (in this case, the annotation `pref`). Similarly, a reading may consist of a witness (for those readings attested in the textual tradition) or of a responsible (`resp`), which expresses the name of the scholar who suggested the preferred reading. The parse tree is visible in Fig. 14. Similar rules have been defined for describing other textual phenomena, such as the degree of likelihood in recovering the original readings, the cause of the variation, editorial interpretations on selected passages and so forth.

The third and last step was to design a listener, a software

```
9  rdgGrp : rdgGrpSep? ana? (rdg)+ |
10           rdgGrpSep? type (rdg)+ ;
11 rdg : (w+)? (wit+|resp+) rdgSep? ;
12 ana : ALPHA_SEQ+ anaSep ;
13 wit : (val)+ ;
14 w : (HEBW|GRCW) ;
15 val :  (WIT_HEBR|WIT_GREEK|WIT_TARG) ;
16
17 rdgSep : RDG_SEP ;
18 rdgGrpSep : RDGGROUP_SEP ;
19 anaSep : DOUBLE_POINT ;
20
21 type : TYPE ;
22 TYPE : 'pref' ;
```

Figure 13. Rules for parsing preferred readings

component which uses the information contained in the CFG to build TEI corresponding elements and attributes. The parser generated by ANTLR is a recursive-descent (or top-down) parser: it starts from the root node of a parse tree and works its way down by vising all the intermediate nodes. Thus, in our example, it starts from the root node `app`, which consists of a location (`loc`) and the lemma (`lem`); then it proceeds further to the location, which in turn entails chapter (`chap`) and verse (`v`) and so forth, up to the token leaf nodes to the extremities of the tree. When visiting a node, the listener executes the desired actions on the node of the tree. Thus, for example, when the listener visits the node `lem`, it performs two tasks: it enters (or discovers) that node and then closes (or finishes) it. When it enters, the opening TEI marker `<lem>` is generated, when it closes the closing marker `</lem>` is generated (see Fig. 15).

IV. Results

After having visited all the nodes of the tree, the parsing system provided by ANTLR generates a TEI compliant XML file, as shown in Fig. 16. All the information contained in the traditional, printed critical apparatus has been successfully parsed and then translated in XML-TEI. All the information concerning witnesses, typology of variation and bibliographical references, moreover, has been extracted and encoded in suitable XML-TEI lists, in order to be linked to the corresponding attribute values. The rigorous and recursive structure of our DSL apparatus has proved to be suitable for a translation to TEI mark-up language. From the encoded text, indeed, it was possible to transform the XML-TEI file, through XSLT style-sheets, back to the printed critical apparatus, without loss of information. The two languages, therefore, are isomorphic. Once the apparatus components have been described and defined, an additional style-sheet is designed in order to generate LaTeX actionable scripts and to get a printed version of both critical text and apparatus, as shown in Fig. 17.[50]

V. Conclusion

The annotation through a DSL is significantly less verbose than the XML-TEI annotation: for instance, the number of characters employed in writing the traditional critical apparatus shown in Fig. 5 is 251 and the TEI counterpart of Fig. 16 is 707. The percentage difference is therefore -64,5%. Carrying out the same calculation on the first three chapters edited so far, the total number of characters of the plain text is 60.844, while the total number of the resulting TEI file is 323.408, with a difference of -81,2%. Compactness is an important feature, especially in case of traditions characterized by a high degree of textual variation, which would require the encoding of long lists of readings and may compromise readability.[51]

Another important advantage is represented by the possibility to establish the set of elements at a later stage. The scholar preparing a digital apparatus through TEI schemas, indeed, must choose from the very beginning which elements are suitable to express his or her interpretation. Interpretation of the semantics of the elements to be encoded and choice of the more appropriate tags to express such an interpretation are simultaneous, coincident activities. On the contrary, the encoding performed through a DSL allows to split the interpretative phase from the operative phase. Being entrusted to the listener, the task of building TEI tags allows to delay such decisions until the end of the whole work-flow. This leads, moreover, to a tighter control on potential semantics errors. It is well known that the TEI's vocabulary makes a large set of markers available for the encoding of textual phenomena which are very similar and often ambiguous. This is the case of elements such as `<q>` and `<quote>`, of attribute such as `@resp` and `@source` or the class of pointers such as `@sameAs`, `@copyOf`, `@corresp` and the like. It may be difficult to decide case by case which one is the most appropriate and to maintain a coherent encoding strategy throughout the study. Ambiguities of this sort may cause an improper use of tags, thus producing semantic errors which can be very difficult to detect, especially in long and complex encoded files. This risk is bypassed in a DSL-based approach. The philologist is exempted from the obligation to decide which strategy is more TEI conformant, and is freed from the cognitive stress due to such a mixture of disciplinary content and cross-disciplinary formalism. Only the first, indeed, is of competence of the scholar, while the second must be addressed only by the digital philologist.

As has been seen in the previous sections, TEI schemas allow a great expressiveness and flexibility in customizing tools of text-critical activity, according to different theoretical

[50] For critical text and apparatus the package LaTeX "eledmac" was used, cf. https://ctan.org/pkg/eledmac.

[51] Cf. the nine principles listed in Lüdeling [58] 488 for the Corpus Encoding Standard (CES) and intended to solve many of the problems of the TEI guidelines mentioned above, in particular the principle of compactness ("markup should be as compact as possible without compromising processability") and readability ("marked up text should still be human readable").

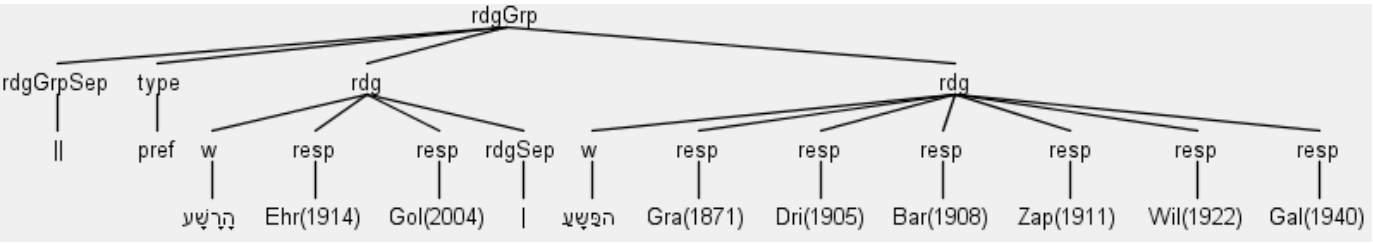

Figure 14. Preferred readings

```
19      public void exitApp(appCriticusParser.AppContext ctx){
20          System.out.print("<app/>");
21      }
22
23          @Override
24      public void enterLemma(appCriticusParser.LemmaContext ctx) {
25              System.out.print("<lem>");
26      }
27          @Override
28          public void exitLemma(appCriticusParser.LemmaContext ctx) {
29      System.out.print("</lem>");
30      System.out.println();
31          }
```

Figure 15. Example of listener's methods in Java code

```
1  <app>
2     <lem wit="#L"><w>הָרַשָּׁע</w></lem>
3     <rdgGrp>
4        <rdg wit="#V"><w>iniquitatem</w></rdg>
5        <rdg wit="#Hy"><w>iniquitas</w></rdg>
6     </rdgGrp>
7     <rdgGrp ana="#subst #sem #n.">
8        <rdg wit="#T"><w>גבראחייבא</w> <w>חייבא</w></rdg>
9        <rdg wit="#Compl. #Ald. #Gra. #Ra."><w>ὁ</w> <w>ἀσεβής</w></rdg>
10       <rdg wit="#P"><w>ܚܛܝܐ</w></rdg>
11       <rdg wit="#AeC #AeB1"><w>ⲘⲚⲞ</w></rdg>
12    </rdgGrp>
13    <rdgGrp ana="#subst #sem #n.">
14       <rdg wit="#GB #GA #GS #Ti."><w>ὁ</w> <w>εὐσεβὴς</w></rdg>
15       <rdg wit="#Syr"><w>ܚܣܝܕܐ</w> <w>ܙܕܝܩܐ</w></rdg>
16       <rdg wit="#Ae"><w>ⲛⲇⲓⲕⲁⲓⲟⲥ</w></rdg>
17       <rdg wit="#An"><w>բարեպաշտ</w></rdg>
18    </rdgGrp>
19    <rdgGrp type="pref">
20       <rdg resp="#Ehr(1914) #Gol(2004)"><w>הָרַשָּׁע</w></rdg>
21       <rdg resp="#Gra(1871) #Dri(1905)
22                  #Bar(1908) #Zap(1911)
23                  #Wil(1922) #Gal(1940)"><w>הַפָּשַׁע</w></rdg>
24    </rdgGrp>
25 </app>
```

Figure 16. TEI compliant apparatus

perceptions. What distinguishes our DSL from XML-TEI is the user-centered approach: by using the DSL, the annotator can avoid TEI technicalities and stay focused on his or her domain-specific research purposes.

The traditional scholar who wishes to prepare a born digital edition, or simply to create a database in order to perform variants database analysis, is not compelled, in this way, to deal with the intricacies of a manual textual encoding. This latter, indeed, is automatically generated by the parser, which falls within the competence of the digital philologist or the computer scientist. After having created the CFG, the parsing results are passed to the computer scientist, who implements the listener, and then to the digital philologist, who knows best how to organize and represent the information according

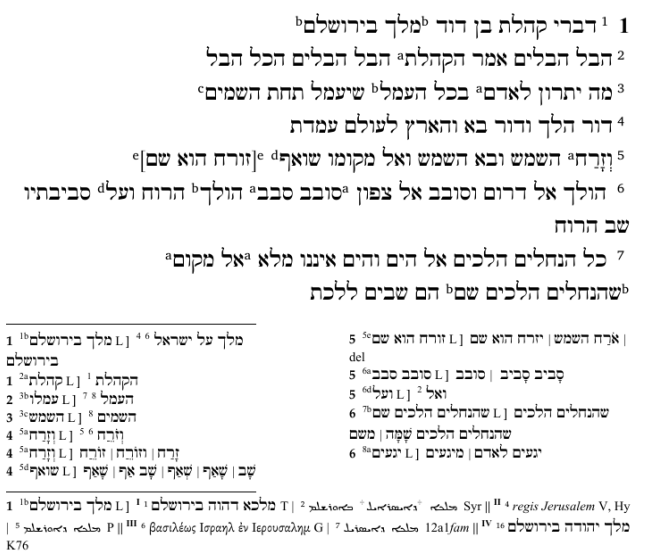

Figure 17. Example of critical edition in LaTeX

to standards. The final results are passed back to the scholar, which has the last word, in order to detect possible errors, inconsistencies or ambiguities. Such an approach, which puts the world of traditional scholarship at its center, may prevent, on one side, traditional scholars with few or no computer skills from straying away from the world of the digital humanities and from the potentialities of computer-assisted text-critical research, and, on the other, domain-specific topics from being addressed by digital philologists or computer scientists with few or no philological expertise.

The widespread suspicion, if not open hostility, against the practices of the digital humanists demonstrated by the traditional philologists arises from the different methodologies and approaches adopted by the respective communities. Digital humanists have defined best practices for the scholarly editing. Unfortunately, these practices can only be adopted with great difficulty by the majority of the traditional academics who, in many cases, likely consider XML based technologies as a barrier, instead of an aid, to their research purposes. For this reason, our domain-centered approach in the development of the supporting technologies is intended to enable, on one side, the traditional philologist to exploit the expressiveness of TEI encoding as an interchange data format and, on the other, to promote the cross-fertilization between the community of

the scholars accustomed to traditional academic methods and the community of the new generation of the digital humanists.

VI. Future Work

Euporia, the web application that hosts the annotation system, is currently just a proof of concept. It needs to be equipped with a text editor that highlights the syntax of the DSLs in use and notifies the syntax errors. Moreover, as in many IDEs, the user should be facilitated by an auto-completion system. Another important difficulty to deal with is represented by the implementation of the listener, which requires to be managed by high skilled programmers. This drawback is bypassed by a general-purpose exporter in XML format that we are releasing. In this way, the computer scientist is exempted from creating TEI compliant XML files, which falls within the competence of the digital philologist, and the digital philologist, in turn, is enabled to reorganize in XML-TEI the relevant information extracted from generic XML documents, through XSLT(S) transformation style-sheets.

References

[1] B. Cerquiglini, *In Praise of the Variant: A Critical History of Philology*. JHU Press, 1999.
[2] G. Mink, "Problems of highly contaminated traditions: the New Testament," in *Studies in Stemmatology*, P. T. v. Reenen and M. v. Mulken, Eds. J. Benjamins Publishing Company, 2004, vol. II, pp. 13–85.
[3] E. J. Epp and G. D. Fee, *Studies in the Theory and Method of New Testament Textual Criticism*. Grand Rapids, Mich.: Wm. B. Eerdmans, 2000.
[4] E. Tov, "Criteria for Evaluating Textual Readings: The Limitations of Textual Rules," *The Harvard Theological Review*, vol. 75, no. 4, pp. 429–448, 1982. [Online]. Available: http://www.jstor.org.emedien.ub.uni-muenchen.de/stable/1509537
[5] ——, *Textual criticism of the Hebrew Bible*, 3rd ed. Minneapolis: Fortress Press, 2012.
[6] C. Segre, *Semiotica filologica — Testo e Modelli Culturali*. Torino: Einaudi, 1979.
[7] H. Fränkel, *Testo critico e critica del testo*, 2nd ed., ser. Bibliotechina del Saggiatore, C. F. Russo, Ed. Firenze: Le Monnier, 1983, no. 31.
[8] D. S. Avalle, *Introduzione alla critica del testo*, ser. Corsi universitari. Torino: G. Giappichelli, 1970.
[9] M. Buzzoni, "Protocol for Scholarly Digital Editions? The Italian Point of View," in *Digital Scholarly Editing: Theories and Practices*, M. J. Driscoll and E. Pierazzo, Eds. Cambridge: Open Book Publishers, 2016, pp. 59–82.
[10] F. Boschetti, *Copisti digitali e Filologi Computazionali*. Roma: CNR Edizioni, 2018.
[11] G. Contini, *Breviario di ecdotica*. Torino: Einaudi, 1990.
[12] M. D. Reeve, *Manuscripts and Methods: Essays on Editing and Trasmission*. Rome: Edizioni di storia e letteratura, 2011.
[13] W. W. Greg, "The Rationale of Copy-Text," *Studies in Bibliography*, vol. 3, pp. 19–36, 1950. [Online]. Available: http://www.jstor.org.emedien.ub.uni-muenchen.de/stable/40381874
[14] D. C. Greetham, *Textual Scholarship — An Introduction*. New York / London: Garland, 1994.
[15] P. Maas, *Textkritik*, 2nd ed. Leipzig: B.G. Teubner, 1950.
[16] S. M. Hockey, *A guide to computer applications in the humanities*. Baltimore ; London : Johns Hopkins University Press, 1983. [Online]. Available: http://archive.org/details/guidetocomputera00hock
[17] R. Pierce, "Multivariate numerical techniques applied to the study of manuscript tradition," in *Tekst Kritisk Teori og Praksis*, B. Fidjestøl, O. E. Haugen, and M. Rindal, Eds., Oslo, 1988, pp. 24–45.
[18] P. T. v. Reenen, M. v. Mulken, and J. Dyk, *Studies in Stemmatology*. John Benjamins Publishing, 1996, vol. I.
[19] P. T. v. Reenen and M. v. Mulken, *Studies in Stemmatology*. J. Benjamins Publishing Company, 2004, vol. II.
[20] G. Pasquali, *Filologia e storia*, ser. Bibliotechina del Saggiatore. Firenze: Le Monnier, 1998, no. 2.
[21] D. Blanke, "Planned Languages — a Survey of some of the main Problems," in *Interlinguistics: Aspects of the Science of Planned Languages*. Walter de Gruyter, Jun. 2011, pp. 63–87.
[22] A. R. Libert, *Artificial Languages*. Oxford University Press, 2018. [Online]. Available: http://oxfordre.com/linguistics/view/10.1093/acrefore/9780199384655.001.0001/acrefore-9780199384655-e-11
[23] D. Blanke, *International Planned Languages*, S. Fiedler and H. Tonkin, Eds. Mondial, 2018.
[24] I. De Vos, C. Macé, and K. Geuten, "Comparing Stemmatological and Phylogenetic Methods to Understand the Transmission History of the 'Florilegium Coislinianum'," in *Ars Edendi Lecture Series vol. II*, ser. Acta Universitatis Stockholmiensis, Studia Latina Stockholmiensia LVIII. Stockholm: Stockholm University Publications, 2012, pp. 107–129.
[25] B. J. P. Salemans, *Building Stemmas with the Computer in a Cladistic, Neo-Lachmannian, Way: The Case of Fourteen Text Versions of Lanseloet Van Denemerken*. University Press, 2000.
[26] F. Ciotti, "La codifica del testo, XML e la Text Encoding Initiative," in *Il Manuale TEI Lite. Introduzione alla Codifica elettronica dei Testi letterari*. Milano: Sylvestre Bonnard, 2005.
[27] M. Burghart, "The TEI Critical Apparatus Toolbox: Empowering Textual Scholars through Display, Control, and Comparison Features," *Journal of the Text Encoding Initiative*, Dec. 2016. [Online]. Available: http://journals.openedition.org/jtei/1520

[28] Rosselli Del Turco, R. et al., "Edition Visualization Technology: A Simple Tool to Visualize TEI-based Digital Editions," *Journal of the Text Encoding Initiative*, 2014. [Online]. Available: http://jtei.revues.org/1077;DOI:10.4000/jtei.1077.

[29] S. Zenzaro, D. Marotta, and A. Bertolacci, "CEED: a Cooperative Web-Based Editor for Critical Editions," in *Settimo Convegno Annuale AIUCD 2018, Book of Abstracts*, Bari, Feb. 2018.

[30] N. Reggiani, *Digital Papyrology: Methods, Tools and Trends*. Berlin / Boston: De Gruyter, 2017.

[31] M. Grassi, C. Morbidoni, M. Nucci, S. Fonda, and F. Di Donato, *Pundit: Creating, exploring and consuming semantic annotations*, Jan. 2013, vol. 1091.

[32] F. Di Donato, C. Morbidoni, S. Fonda, A. Piccioli, M. Grassi, and M. Nucci, *Semantic annotation with Pundit: A case study and a practical demonstration*, Sep. 2013.

[33] S. Timpanaro, *The Genesis of Lachmann's Method*. Chicago / London: University of Chicago Press, 2005.

[34] E. J. Kenney, G. Ravenna, and A. Lunelli, *Testo e metodo : aspetti dell'edizione dei classici latini e greci nell'età del libro a stampa*. Roma: Gruppo editoriale internazionale, 1995.

[35] Y. G. Gane, "Apparatp critico," in *Dizionario della Terminologia filologica*. Accademia University Press, 2013.

[36] B. M. Metzger and B. Ehrman, *The Text of New Testament — Its Transmission, Corruption, and Restoration*. New Tork/Oxford: Oxford Univ. Press, 2005. [Online]. Available: http://archive.org/details/TheTextOfNewTestament4thEdit

[37] V. N. Grishin, "Formalized language," in *Encyclopaedia of Mathematics*. Kluwer Academic Publishers, 1989, vol. 4, pp. 61–62, google-Books-ID: s9F71NJxwzoC.

[38] M. Fowler, *Domain-Specific Languages*. Pearson Education, 2010.

[39] T. Parr, *The Definitive ANTLR 4 Reference*. Pragmatic Bookshelf, 2012.

[40] ——, *Language Implementation Patterns: Create Your Own Domain-specific and General Programming Languages*. Pragmatic Bookshelf, 2010.

[41] W. Rudolph, K. Elliger, A. Alt, O. Eissfeldt, P. Kahle, R. Kittel, H. P. Rüger, and G. E. Weil, *Biblia Hebraica Stuttgartensia*. Stuttgart: Deutsche Bibelgesellschaft, 1997.

[42] G. V. D. Schenker, A. Shenker, Y. A. P. Goldman, G. J. Norton, and A. V. D. Kooji, *Biblia Hebraica Quinta: Megilloth: Ruth, Canticles, Qoheleth, Lamentations, Esther*, bilingual edition ed. Stuttgart: Deutsche Bibelgesellschaft, Jan. 2004.

[43] E. Tov, "Electronic Scripture Editions (With an Appendix Listing Electronic Editions on the Internet [2014])," in *The Text of the Hebrew Bible and Its Editions: Studies in Celebration of the Fifth Centennial of the Complutensian Polyglot*, A. P. Otero and P. A. T. Morales, Eds. Leiden / Boston: Brill, 2016.

[44] ——, "Computer-assisted tools for textual criticism," *Tradition and Innovation in Biblical Interpretation*, 2011.

[45] E. Tigchelaar, "Editing the Hebrew Bible: An Overview of some Problems," in *Editing the Bible: Assessing the Task Past and Present*, J. S. Kloppenborg, S. John, and H. N. Judith, Eds. Atlanta: Society of Biblical Literature, 2012.

[46] E. Tov, "Hebrew Scripture Editions: Philosophy and Praxis," in *Hebrew Bible, Greek Bible, and Qumran - Collected Essays*, ser. Texts and Studies in Ancient Judaism. Tübingen: Mohr Siebeck, 2008, no. 121.

[47] R. Hendel, "The Oxford Hebrew Bible: Prologue to a New Critical Edition," *Vetus Testamentum*, vol. 58, pp. 324–351, 2008.

[48] M. Segal, "Methodological Considerations in the Preparation of an Edition of the Hebrew Bible," in *The Text of the Hebrew Bible and Its Editions*. Leiden, The Netherlands: Interactive Factory, 2017.

[49] R. Hendel, "From Polyglot to Hypertext," in *The Text of the Hebrew Bible and Its Editions*. Leiden, The Netherlands: Interactive Factory, 2017.

[50] G. Mugelli, F. Boschetti, R. Del Gratta, A. M. Del Grosso, F. Khan, and A. Taddei, "A user-centred design to annotate ritual facts in ancient greek tragedies," *BICS*, vol. 59, 2, pp. 103–120, 2016.

[51] D. Barthélemy, *Critique textuelle de l'Ancien Testament*. Fribourg: Academic Press, 2015, vol. 5.

[52] S. R. Driver, "Ecclesiastes," in *Biblia hebraica*, R. Kittel, Ed. Lipsiae: Hinrichs, 1905.

[53] P. G. Borbone, "Orientamenti Attuali dell'Ecdotica della Bibbia Ebraica: Due Progetti di Edizione dell'Antico Testamento Ebraico," *Materia Giudaica*, vol. 6, no. 1, pp. 28–35, 2001.

[54] J. Cummings, "The Text Encoding Initiative and the Study of Literature," in *A Companion to Digital Literary Studies*, R. Siemens and S. Schreibman, Eds. John Wiley & Sons, 2013, pp. 451–76.

[55] W. W. Greg, *The Calculus of Variants — An Essay on Textual Criticism*. Oxford: Clarendon, 1927.

[56] A. Catastini, *L'itinerario di Giuseppe: studio sulla tradizione di Genesi 37-50*. Dipartimento di Studi Orientali. Studi Semitici, N. S. 13; Roma: Università degli studi "La Sapienza", 1995.

[57] ——, *Isaia ed Ezechia — Studio di storia della tradizione di II Re 18-20 // Is. 36-39*. Dipartimento di Studi Orientali. Studi Semitici, N. S. 6; Roma: Università degli studi "La Sapienza", 1989.

[58] A. Lüdeling and M. Kytö, *Corpus linguistics: An international Handbook*. W. de Gruyter, 2008, google-Books-ID: EXIoAQAAIAAJ.