

# ekdosis: Using LuaLaTeX for Producing TEI xml Compliant Critical Editions and Highlighting Parallel Writings

Robert Alessi

## ► To cite this version:

Robert Alessi. ekdosis: Using LuaLaTeX for Producing TEI xml Compliant Critical Editions and Highlighting Parallel Writings. *Journal of Data Mining and Digital Humanities*, Episciences.org, 2020, Special Issue on Collecting, Preserving, and Disseminating Endangered Cultural Heritage for New Understandings through Multilingual Approaches. hal-02779803v3

HAL Id: hal-02779803

<https://hal.archives-ouvertes.fr/hal-02779803v3>

Submitted on 31 Oct 2020

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# ekdosis: Using Lua $\LaTeX$ for Producing TEI $\text{xml}$ Compliant Critical Editions and Highlighting Parallel Writings

Robert Alessi<sup>1</sup>

<sup>1</sup>CNRS UMR Unit 8167 “Orient & Méditerranée”, Sorbonne–Université, France

## Abstract

ekdosis is a Lua $\LaTeX$  package written by R. Alessi designed for multilingual critical editions. It can be used to typeset texts and different layers of critical notes in any direction accepted by Lua $\TeX$ . Texts can be arranged in running paragraphs or on facing pages, in any number of columns which in turn can be synchronized or not. Database-driven encoding under  $\LaTeX$  allows extraction of texts entered segment by segment according to various criteria: main edited text, variant readings, translations or annotated borrowings between texts. In addition to printed texts, ekdosis can convert  $\text{.tex}$  source files so as to produce TEI  $\text{xml}$  compliant critical editions. It will be published under the terms of the GNU General Public License (GPL) version 3.

## Keywords

$\LaTeX$ ; Lua; TEI  $\text{xml}$ ; multilingual critical editions; alignment; segmentation

## I INTRODUCTION

The name of the software that is presented here, ekdosis, derives from a Greek action noun—ἐκδοσις—the meaning of which is: “publishing a book”, and also in concrete sense: “a publication, treatise”. For us moderns, this term refers to a long tradition of scholarly work consisting in establishing from manuscript evidence the texts of Greek and Latin classics that were handed down through the Middle Ages to the time of the first printed editions. It is not quite sufficient to mention that no original of such text survives: as Jean Irigoin vividly pointed out in the inaugural lecture that he delivered at the Collège de France in 1986, Sophocles witnessed the building of the Parthenon. As a result, what still can be seen or read from the remains preserved on the Acropolis, at the Louvre or the British Museum, is pretty much the same as Sophocles already could see or read in his time. But the modern reader of the same Sophocles’ preserved tragedies has to rely on manuscripts the most ancient of which was copied over a millennium and a half after Sophocles died.<sup>1</sup> Scientific textual criticism gradually emerged from the last years of the xvii<sup>th</sup> century out of the rejection of the idea that any improvement of classical texts should be conducted on the basis of vulgate texts transmitted in printed books.<sup>2</sup> It must be noted that the first impulse came from the *Textus receptus* of the New Testament. Not only scholars had to deal with a very large number of versions and scholia and countless variant readings, but in many aspects, they were facing unprecedented problems posed by the existence of early patristic citations and ancient translations in various languages, some of which even showed that certain passages of the received Greek text had been deliberately falsified.

---

<sup>1</sup>Irigoin [1997, p. 7].

<sup>2</sup>On this movement, see Irigoin [1997, pp. 3–27] and Reynolds and Wilson [1991, pp. 207–41].

Contrary to what one might think, the foregoing remarks are far from being unrelated to digital humanities. Whether they are printed or digital, modern critical editions always exhibit reconstructed texts, and both convey the legacy of those early scholarly attainments. The texts reconstructed come either from manuscripts under the title of the edited work (direct tradition) or from evidence of explicit citations or parallel passages preserved in other authors or from translations in other languages made on the basis of manuscripts which may or may not have survived (indirect tradition). Both kind of editions come equipped with an apparatus criticus in which is mentioned all the evidence that was used to build the edited text. However, this is where the two kinds of editions—printed or digital—part ways. Surely, reading a traditional and well formed apparatus criticus requires a specific training and is only meant for experienced readers: all notes are grouped together in the form of a paragraph of its own, freely composed in Latin at least when it comes to editing a classical text. The conventions are many, whether they concern Latin technical abbreviated terms, using lower-case Greek letters to indicate intermediate lost manuscripts and upper-case Roman letters to indicate preserved manuscripts, using spaces for grouping the sigla into families and exponents for pointing to traceable corrections, where, to take an example, number 1 always refers to first-hand corrections, number 2 to a second hand and so forth. Another source of difficulty arises from the style of the apparatus: some are set out fully and display the words of the adopted text and the part of the tradition they belong to as well as the variant readings, while others only mention the deviant readings which are rejected by the editor.

Arguably, getting oneself familiarized with these conventional rules is not unrelated to learning a language equipped with terms, grammar rules and style embellishments. It came into existence out of over three centuries of tradition and cultural facts and is immediately accessible to human mind's natural ability to use language and interpret conventional symbols. It may be true that reading an apparatus criticus requires training and effort. Nevertheless, this task is part of the curriculum and remains the natural, traditional way to go. But it is quite inaccessible to a computer, unless every item of information has been encoded in the rather dumb format that is suited to machines.

On the other hand, critical editions in print have their own limitations. For example, editors of classical texts are used to save space by not reporting trivial mistakes committed by medieval scribes—and rightfully so. However, as for ancient authors who wrote in subdialects, the manuscripts may display forms that are not considered acceptable by modern linguistics even though such forms may go back to the archetype of the preserved witnesses. The question therefore arises as to which text was in circulation in ancient times. We may have grounds to normalize these texts in modern critical editions. But do we have the right to remove those allegedly 'faulty' variants from the apparatus to save space and ease reading, thus making them disappear from the tradition for ever?

Another limitation of editions in print comes from "fluid forms of transmission", as in the case of technical and popular literature.<sup>3</sup> To take here just one example, we have a large amount of works from Late Antiquity and the Middle Ages in which the primary intention of the scholars was to provide useful, practical or scientific information. Depending on the intellectual *milieu* or geographic area in which such texts were actually circulating, they were adapted to the needs of students, namely expanded, abridged or rephrased. Their content was also constantly updated as science progressed, including by their original authors themselves, which could lead to the existence of several recensions of the text in their lifetime. It goes without saying that putting in print such texts—and which ones exactly?—may become quite an intricate business for it is much more about circulation and method of redaction than archetype and authorship.

---

<sup>3</sup>On this concept and for actual examples of such texts and authors, see Reynolds and Wilson [1991, pp. 234–7].

One final limitation of editions in print is about indirect tradition of transmitted texts.<sup>4</sup> Strictly speaking, references to texts quoted by the author of the edited text do not constitute indirect tradition. But they may be collected into the critical notes as an additional layer called *apparatus fontium*. As for references to the edited text by other authors, they are collected in what is called an *apparatus testium*. It may happen that such quotations attest variant readings from the manuscripts of the direct tradition or even provide new variants of their own. These *testimonia* can be found in the same language as the edited text or in another language. To take a common example of such phenomena, many Greek books of technical literature have been translated into Syriac and/or Arabic throughout Late Antiquity and the Middle Ages. Out of necessity, but also for obvious reasons of space available on the pages, editions in print give no access to indirect tradition or translations of the main text in other languages. After all, only deviant readings are of interest to establish the edited text.

The various limitations described above are damaging in more than one respect. First, every detail that the editor decided to discard, regardless of its relevance to the purpose of the edition, is lost permanently as in the case of dialectal coloring of ancient books. Second, passages collected as indirect tradition are only available as references in the *apparatus testium*. This may be acceptable for short passages quoted by other authors who wrote in the same language as the primary author. However, as translations cannot be compared to the original text, the reader is refrained from bestowing attention upon major ones to understand better difficult passages. But there is more to say in this respect. As a matter of fact, it is only natural that in many cases only original texts are considered worthy of interest. In comparison, translations are looked at as satellite texts. As a result, unless they are made by someone considered a prominent author in a different cultural area, translations are likely to gradually lose interest.

To conclude on these issues, print publications and digital editions are often contrasted as they belonged to two different worlds.<sup>5</sup> It is commonly said that the content of editions in print is the result of the binding of the book itself as an object, whereas digital editions, in which format and presentation are by definition separated from content, are free from limitations coming from such bindings. To sum up from the foregoing considerations, this statement is likely to be qualified: as already seen above, the apparatus criticus must be looked at as a brilliant production of mind refined by centuries of scholarly tradition—and surely tradition must go on—arguably not as compact paragraphs that require special and painful training to be ‘decoded’. On the other hand, what editions in print do not provide are what D.J. Mastronarde and R.J. Tarrant have called “actionable texts for use in digital research”,<sup>6</sup> namely database-driven texts allowing the reader to select annotations and display or arrange translations, parallel passages or borrowings in a variety of ways.

ekdosis can be seen as an attempt at combining the two approaches as will be illustrated by the following.

## II USED PROGRAMMING LANGUAGES

At the time of writing, v1.0 of ekdosis, which is about to be released as a Lua<sup>L</sup>A<sup>T</sup>E<sup>X</sup> package under the terms of the GNU General Public License (GPL) version 3, is written mostly in T<sup>E</sup>X and Lua (see fig. 1 for details).<sup>7</sup> Both languages are used to produce PDF output, while Lua is almost exclusively used to produce at the same time T<sup>E</sup>X `xml` output. The concern to follow the

---

<sup>4</sup>See above p. 2.

<sup>5</sup>For a good illustration of this point, see Digital Latin Library [nodate, “Content, not Display”].

<sup>6</sup>Mastronarde and Tarrant.

<sup>7</sup>Additionally, an AUCT<sup>E</sup>X Lisp style file is provided to perform code-folding in the emacs text editor. AUCT<sup>E</sup>X is an extension to emacs for writing and formatting T<sup>E</sup>X files. See <https://www.gnu.org/software/auctex>.

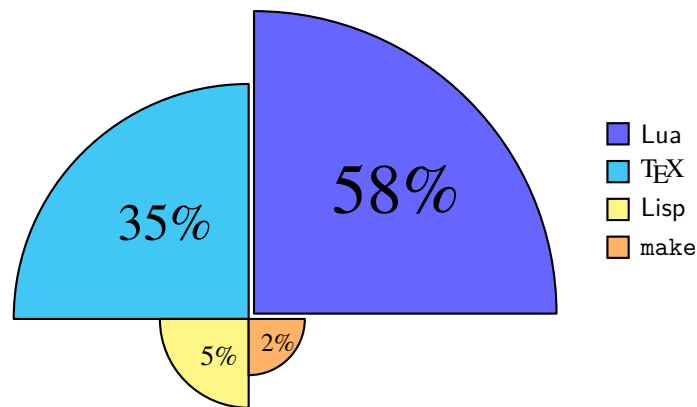


Figure 1: Programming languages used in ekdosis

tradition established by great printed editions explains why L<sup>A</sup>T<sub>E</sub>X has been primarily preferred as a typesetting system with the intent of reviving a tradition, now almost forgotten, in which the brilliance of the presentation was not dissociated from the quality of the academic work. It is worth recalling here that both T<sub>E</sub>X and Lua come from an academic source. Besides, L<sup>A</sup>T<sub>E</sub>X, contrary to xml, is a natural language designed for use in typesetting complex documents, including as many languages one wishes to have printed in any writing direction, such as critical editions where footnotes and other kind of annotations can be particularly abundant. Document processing through compilation is also an argument in favor of L<sup>A</sup>T<sub>E</sub>X, let alone, as has just been said, that it is meant to serve complexity: in addition to fine PDF output, compilation can produce with no additional effort, from a single source file, various types of outputs, such as OpenDocument, html or TEI xml files.

By using Lua as an additional scripting language, notably for pattern matching and string handling functions, any portion of the code, as it is compiled by L<sup>A</sup>T<sub>E</sub>X, can be intercepted and passed on to Lua functions for further processing and then returned to L<sup>A</sup>T<sub>E</sub>X. Some functions are designed to support document processing through L<sup>A</sup>T<sub>E</sub>X while others act as a “L<sup>A</sup>T<sub>E</sub>X to TEI xml” converter as will be demonstrated shortly by several examples. The main features of the software are as follows:

1. Multilingual critical editions:— ekdosis can be used to typeset texts and various layers of critical notes<sup>8</sup> in any direction accepted by LuaT<sub>E</sub>X, which makes it adapted to rare or lesser-known languages. Texts can be arranged in running paragraphs or on facing pages, in any number of columns which in turn can be synchronized or not. It is also suitable for complex layouts, such as Arabic poetry or images where three-way alignment is required,<sup>9</sup> diagrams, etc.
2. Database-driven encoding under L<sup>A</sup>T<sub>E</sub>X, which allows extractions of texts entered segment by segment according to various criteria: main edited text, variant readings or translated texts and annotated borrowings between texts. From a given main text—whether it be critically edited or not—ekdosis can select, display and arrange in aligned columns other recensions or parallel texts.
3. Academic background:— ekdosis comes in support of a seminar to be held in fall 2020

<sup>8</sup>More on this technical point below p. 9.

<sup>9</sup>On this technique, see <https://tei-c.org/release/doc/tei-p5-doc/en/html/SA.html#SACSXA>. Three-way alignment is an extension of the TEI linkGrp mechanism (<https://tei-c.org/release/doc/tei-p5-doc/en/html/SA.html#SACSAL>) which uses *Scalable Vector Graphics* (SVG) notation and is already provided by ekdosis as the examples to come will show.

onwards about the scientific controversies raised by the many discussions about medical education which took place in the Greek, Arabic and Byzantine worlds.<sup>10</sup> The texts to be discussed in the seminar are characterized by their complexity, as some of them are made out of several recensions, while most of them feature borrowings—either explicit or not—translations (from Greek into Arabic or vice versa), parallel writings and commentaries.

### III MAIN FEATURES

These features will be illustrated here through an excerpt from the Latin text of Caesar's *Gallic War* (Book VI, 13.1) as it is read in the French edition of the *Budé* series.<sup>11</sup>

As can be seen from fig. 2, as lines are not numbered, the notes in the apparatus refer to the sectional divisions of the edited text. That said, the apparatus criticus can be readily scrutinized in light of the

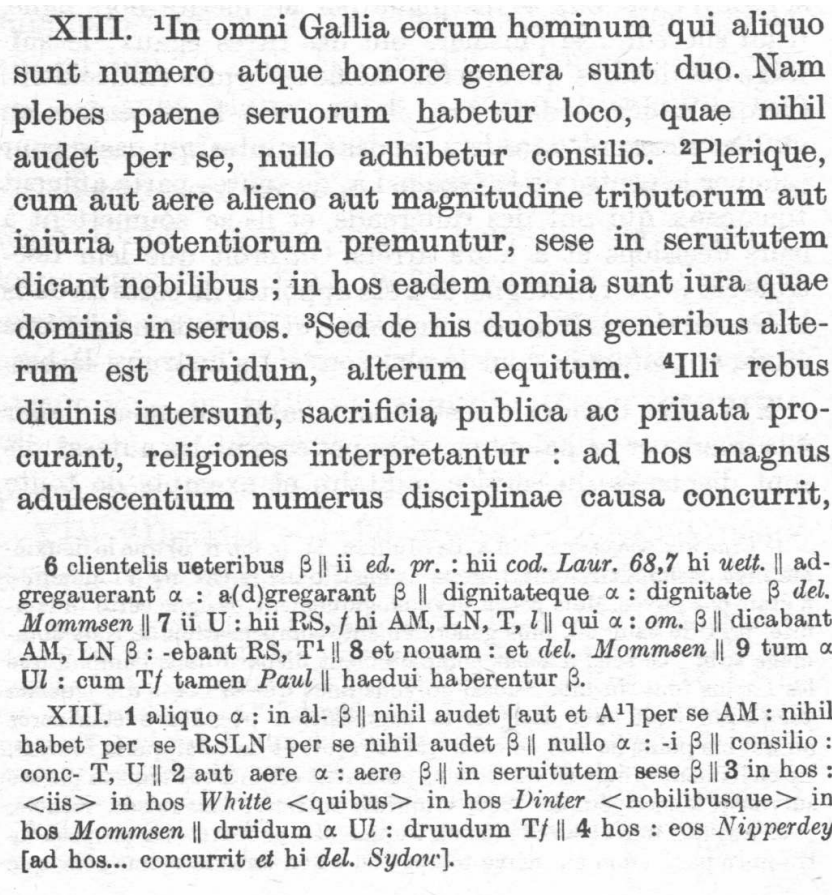


Figure 2: Caesar, *Gallic War*, VI.13 (ed. Constans)

rules outlined above pp. 1–2. For instance, section XIII.1 shows examples of abbreviated readings and sigla, either ‘grouped’ or not, some of which refer to families of manuscripts (in lower-case Greek letters) while others to single witnesses (in upper-case Roman letters).

First, ekdosis is loaded in the preamble and instructed to output both PDF and TEI xml files at the same time, like so:—

```
11 \usepackage[telexport=tidy]{ekdosis}
```

<sup>10</sup>Joint project between the French CNRS and the German BAW: R. Alessi (CNRS, Paris, France) and I. Grimm-Stadelmann (Bayerische Akademie der Wissenschaften, Munich, Germany).

<sup>11</sup>Caesar, VI, 13.1, p. 185.

Then, in a way which is very close to TEI xml encoding, witnesses must be ‘declared’ in the preamble of the  $\LaTeX$  file:<sup>12</sup>—

```

27 \DeclareWitness{A}{A}{\emph{Bongarsianus} 81}[
28     msName=\emph{Bongarsianus},
29     settlement=Amsterdam,
30     idno=81,
31     repository=University Library,
32     origDate=s. IX--X]
33 \DeclareHand{A1}{A}{A\textsuperscript{1}}[\emph{Emendationes
34     scribae ipsius}]

```

As can be seen, `\DeclareWitness` requires three mandatory arguments enclosed between curly braces used to specify consecutively: the identifier of the witness to be used as `xml:id`,<sup>13</sup> the rendition of the siglum to be used in the apparatus criticus and a short description used to build the *conspectus siglorum* to be printed at the forefront of the edited text.<sup>14</sup> Finally, other items of information can be specified in a further optional argument enclosed between square brackets. The three mandatory arguments of `\DeclareHand` specify in turn: the unique `xml:id` to be used, the `xml:id` of the witness the hand is related to and lastly the rendition to be printed in the apparatus criticus. Finally, *families* of witnesses can be declared as shorthands:<sup>15</sup>—

```

56 \DeclareShorthand{a}{\alpha}{A,M,B,R,S,L,N}
57 \DeclareShorthand{b}{\beta}{T,f,U,l}

```

Where the first argument of `\DeclareShortHand` is the `xml:id`, the second its rendition to be used in the apparatus criticus and the third a comma-separated list of declared witnesses.

Once witnesses have been declared, Caesar’s *Gallic War*, Book VI, XIII.1 authoritative text composed of lemmata can be encoded with *ekdosis* in a way that is very close to TEI xml encoding:—

---

**Listing III.1** Caesar, *BG*, VI, 13.1:  $\LaTeX$  source code

---

```

85 \begin{latin}
86   \ekddiv{head=XIII, depth=2, n=6.13, type=section}
87   \begin{segment}
88     In omni Gallia eorum hominum qui \app{
89       \lem[wit=a]{aliquo}
90       \rdg[wit=b, alt=in al-]{in aliquo}}
91     sunt numero atque honore genera sunt duo. Nam plebes paene
92     seruorum habetur loco, quae \app{
93       \lem[wit={A,M}, alt={nihil audet (aut et \getsiglum{A1})}
94       per se]}{nihil audet per se}
95       \rdg[wit=A1,nordg]{nihil aut et per se}
96       \rdg[wit={R,S,L,N}]{nihil habet per se}
97       \rdg[wit=b]{per se nihil audet}}, \app{
98       \lem[wit=a]{nullo}
99       \rdg[wit=b]{nulli}} adhibetur \app{
100      \lem{consilio}

```

<sup>12</sup>Line numbers refer to the file used to build the examples.

<sup>13</sup>*ekdosis* checks that any declared `xml:id` is unique before registering it.

<sup>14</sup>See below, fig. 4.

<sup>15</sup>Another approach which will not be described here for the sake of simplicity is to assign the `xml:id` of the family as the container of the individual witnesses that are part of the family.

1 XIII. In omni Gallia eorum hominum 2 qui aliquo sunt numero atque honore ge- 3 nera sunt duo. Nam plebes paene seruo- 4 rum habetur loco, quae nihil audet per se, 5 nullo adhibetur consilio.	XIII. Throughout all Gaul there are two orders of those men who are of any clas rank and dignity: for the commonality is con: held almost in the condition of slaves, and ne s dares to undertake nothing of itself, and escl is admitted to no deliberation. initi
--	--

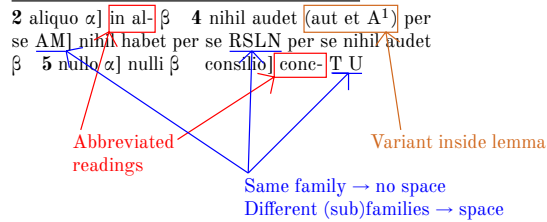


Figure 3: Apparatus criticus generated by ekdosis

```

101 \rdg[wit={T, U}, alt=conc-]{concilio}}.
102 \end{segment}
103 \end{latin}

```

Close examination of the two commands `\lem` and `\rdg` that were used in listing III.1 shows that while some optional ‘name-value’ arguments such as `wit` have TEI equivalent attributes, others do not, as `alt` (listing III.1, ll. 90, 93–4 and 101) and `nordg` (l. 95).<sup>16</sup> The principle is that ekdosis uses some of them for PDF or TEI output only, and others for both outputs. Obviously, `wit` is one of the latter, although values are rendered differently in PDF and in TEI. On the other hand, `alt` is used to introduce an alternate way of inserting words in the apparatus criticus in print, while keeping safe what is to be found in TEI output. Moreover, as can be seen on lines 93–4, `alt` has been used to insert a subvariant in the lemma part, as a consequence of which `nordg` was required (l. 95) to remove from PDF output words that would have been otherwise redundant.

Fig. 3 shows the apparatus criticus generated from the source code given above. The corresponding TEI xml output follows:—

---

**Listing III.2** Caesar, *BG*, VI, 13.1: TEI output

---

```

178 <div xml:id="div-latin_1" xml:lang="la">
179   <div type="section" n="6.13">
180     <head>XIII</head>
181     <p>
182       <seg xml:id="latin_1">In omni Gallia eorum hominum qui
183       <app>
184         <lem wit="#A #M #B #R #S #L #N">aliquo</lem>
185         <rdg wit="#T #f #U #l">in aliquo</rdg>
186       </app>sunt numero atque honore genera sunt duo. Nam
187       plebes paene seruorum habetur loco, quae
188       <app>
189         <lem wit="#A #M">nihil audet per se</lem>
190         <rdg wit="#A1">nihil aut et per se</rdg>
191         <rdg wit="#R #S #L #N">nihil habet per se</rdg>
192         <rdg wit="#T #f #U #l">per se nihil audet</rdg>
193       </app>,

```

<sup>16</sup>The current version of ekdosis provides about ten such optional ‘named’ arguments.



```

194     <app>
195         <lem wit="#A #M #B #R #S #L #N">nullo</lem>
196         <rdg wit="#T #f #U #l">>nulli</rdg>
197     </app>adhibetur
198     <app>
199         <lem>consilio</lem>
200         <rdg wit="#T #U">concilio</rdg>
201     </app>.</seg>
202 </p>
203 </div>
204 </div>

```

---

#### IV ALIGNMENT AND SEGMENTATION

An alert reader will have noticed from listing III.2 two `<div>` elements at lines 178–9. As a matter of fact, lines 178 and 204 from the TEI output come from `\begin{latin} ... \end{latin}` from the corresponding L<sup>A</sup>T<sub>E</sub>X source file (listing III.1, lines 85 and 103), while TEI lines 179, 180 and 203 are generated by one single command provided by ekdosis, `\ekddiv` (listing III.1, l. 86). Moreover, `\begin{segment} ... \end{segment}` (listing III.1, lines 87 and 102) has been converted into `<seg>` elements by ekdosis (listing III.2, lines 182 and 201).

These features call for two remarks. First, ekdosis knows where any opened TEI element that is allowed to nest recursively, such as `<div>`, `<lg>` and the like, is to be closed, even though, as in the case of the `\ekddiv` command, there is no explicit indication of the point where the closure occurs. Thoroughly scanning L<sup>A</sup>T<sub>E</sub>X source files with Lua functions which involve complex string matching, reverse string matching and recursions was required, as L<sup>A</sup>T<sub>E</sub>X ‘open’ commands such as `\chapter` or `\section` only act as milestones, contrary to nested TEI elements. ekdosis converts these commands into TEI ‘numbered’ textual divisions, namely `<div1>` to `<div7>`. Moreover, in addition to L<sup>A</sup>T<sub>E</sub>X standard textual divisions, `\ekddiv` was needed to meet the requirements of classical and literary texts the divisions of which depend on many different received traditions. As can be seen from listing III.1, line 86, `\ekddiv` processes a comma-separated list of ‘name-value’ arguments. Some, as head, correspond to TEI subsequent elements, while others correspond to attributes of `<div>` elements. As to `depth`, which has no TEI equivalent, ekdosis uses its value to build un-numbered TEI divisions allowed to nest recursively—or not allowed to—in accordance with their declared hierarchic depth. This mechanism gives the flexibility that is needed in printed editions of classical texts. Finally, it must be noted that format and presentation have been carefully separated: in the PDF output (fig. 3), chap. XIII, which is printed at the beginning of the current paragraph in Roman capital letters followed by a dot and an en quad is merely encoded in `\ekddiv` as `head=XIII` with no dot. Other commands, which are not discussed here, allow to set the format of any textual division that is used.

The second remark is about the two environments `\begin{latin} ... \end{latin}` and `\begin{segment} ... \end{segment}` (listing III.1, ll. 85, 87, 102 and 103): as can be seen from the TEI output (listing III.2, ll. 178 and 182), the corresponding `<div>` and `<seg>` elements have been given `xml:id` attributes by ekdosis. This important feature allows for alignment of parallel texts from multilingual corpora. To return to the example of Caesar’s *Gallic War* VI, 13.1 presented here, the alignment has been set in the preamble of the L<sup>A</sup>T<sub>E</sub>X source file like so:—

---

**Listing IV.1** Alignment setting in ekdosis

---

```
14 \SetEkdosisAlignment{
15   tcols=3,
16   lcols=3,
17   texts=latin[xml:lang="la"];
18         english[xml:lang="en"];
19         french[xml:lang="fr"],
20   apparatus=latin,
21   segmentation=auto}
```

---

It is worth mentioning that `\SetEkdosisAlignment` provides further options suited for corpora in which translations or parallel passages are sufficient in number to be provided in separate files.<sup>17</sup> However, this point will not be discussed here, as representing all the possible contrivances leads to much complexity and would be too long to consider.

In listing IV.1, `tcols` and `lcols` stand for “total number of columns” and “columns to be printed on the left-hand page”<sup>18</sup> respectively. The next option, `texts`, defines names of environments that are to receive texts to be aligned, namely: `latin`, `english` and `french`—further sub-options can be specified between square brackets, such as `xml:lang` attributes. Then, the `apparatus` option, just as `texts`, takes a semicolon-separated list of previously defined environments that shall receive at least one layer of apparatus criticus. As already said (see above p. 2 and point 1 p. 4), several layers of critical notes can be defined.<sup>19</sup> Finally, `segmentation=auto` instructs `ekdosis` to automatically increment the `xml:id` attributes associated to each segment of text delimited by the  $\LaTeX$  environment `\begin{segment} . . . \end{segment}` as in listing III.1, ll. 87 and 102.

The complete  $\LaTeX$  body text that was used to build Caesar’s *Gallic War*, VI, 13.1 follows:—

---

**Listing IV.2** Caesar’s *BG*, VI, 13.1: complete alignment environment

---

```
84 \begin{alignment}
85   \begin{latin}
86     \ekddiv{head=XIII, depth=2, n=6.13, type=section}
87     \begin{segment}
88       In omni Gallia eorum hominum qui \app{
89         \lem[wit=a]{aliquo}
90         \rdg[wit=b, alt=in al-]{in aliquo}}
91       sunt numero atque honore genera sunt duo. Nam plebes paene
92       seruorum habetur loco, quae \app{
93         \lem[wit={A,M}, alt={nihil audet (aut et \getsiglum{A1})
94         per se}]{nihil audet per se}
95         \rdg[wit=A1,nordg]{nihil aut et per se}
96         \rdg[wit={R,S,L,N}]{nihil habet per se}
97         \rdg[wit=b]{per se nihil audet}}, \app{
```

---

<sup>17</sup>In such cases, it is possible to instruct `ekdosis` not to produce PDF output or to select which parallel passages are to be printed. See above point 2 p. 4.

<sup>18</sup>That is the number of columns to be printed on the left-hand page, out of the total number of columns previously defined. For example, `tcols=2` and `lcols=1` will print two aligned texts on facing pages.

<sup>19</sup>This is beyond the scope of this paper. `ekdosis` sets one layer by default and provides a command, `\DeclareApparatus`, that allows to define further types of notes. Additional types of notes can then be accessed by optional arguments such as `\app[type=<value>]{}` or `\note[type=<value>]{}`.

```

98     \lem[wit=a]{nullo}
99     \rdg[wit=b]{nulli}} adhibetur \app{
100     \lem{consilio}
101     \rdg[wit={T, U}, alt=conc-]{concilio}}.
102     \end{segment}
103 \end{latin}
104 \begin{english}
105     \ekddiv{head=XIII, depth=2, n=6.13, type=section}
106     \begin{segment}
107         Throughout all Gaul there are two orders of those men who are of
108         any rank and dignity: for the commonality is held almost in the
109         condition of slaves, and dares to undertake nothing of itself,
110         and is admitted to no deliberation.
111     \end{segment}
112 \end{english}
113 \begin{french}
114     \ekddiv{head=XIII, depth=2, n=6.13, type=section}
115     \begin{segment}
116         Partout en Gaule il y a deux classes d'hommes qui comptent et
117         qui sont considérés. Quant aux gens du peuple, ils ne sont guère
118         traités autrement que des esclaves, ne pouvant se permettre
119         aucune initiative, n'étant consultés sur rien.
120     \end{segment}
121 \end{french}
122 \end{alignment}

```

Aside from the technicalities inherent in any critically edited text where variant readings are abundant in number, it must be noted that the three environments `latin`, `english` and `french` defined with `\SetEkdosisAlignment` have been used in an easy and uncomplicated manner. Then, inside each environment, passages that correspond to each other have been inserted inside a `segment` environment which is provided by `ekdosis`. As segmentation has been set to `auto`, `segment` environments must be equal in number inside each portion of parallel or translated text.<sup>20</sup> Once it is compiled by  $\text{\LaTeX}$ , the source file from listing IV.2 produces the following TEI output:—

---

**Listing IV.3** Caesar's *BG*, VI, 13.1: complete TEI alignment

---

```

176 <text>
177   <body>
178     <div xml:id="div-latin_1" xml:lang="la">
179       <div type="section" n="6.13">
180         <head>XIII</head>
181         <p>
182           <seg xml:id="latin_1">In omni Gallia eorum hominum qui
183             <app>
184               <lem wit="#A #M #B #R #S #L #N">aliquo</lem>
185               <rdg wit="#T #f #U #l">in aliquo</rdg>
186             </app>sunt numero atque honore genera sunt duo. Nam

```

---

<sup>20</sup>If segmentation had been set to `noauto`, it would have been required to give each `segment` environment manually inserted marks, like so: `\begin{segment}[xmlid=lat1, corresp={en1, fr1}]`. `noauto` provides flexibility needed when not every piece of text has matches in other defined domains.

```

187     plebes paene seruorum habetur loco, quae
188     <app>
189         <lem wit="#A #M">nihil audet per se</lem>
190         <rdg wit="#A1">nihil aut et per se</rdg>
191         <rdg wit="#R #S #L #N">nihil habet per se</rdg>
192         <rdg wit="#T #f #U #l">per se nihil audet</rdg>
193     </app>,
194     <app>
195         <lem wit="#A #M #B #R #S #L #N">nullo</lem>
196         <rdg wit="#T #f #U #l">>nulli</rdg>
197     </app>adhibetur
198     <app>
199         <lem>consilio</lem>
200         <rdg wit="#T #U">concilio</rdg>
201     </app>.</seg>
202 </p>
203 </div>
204 </div>
205 <div xml:id="div-english_1" xml:lang="en">
206     <div type="section" n="6.13">
207         <head>XIII</head>
208         <p>
209             <seg xml:id="english_1">Throughout all Gaul there are two
210             orders of those men who are of any rank and dignity: for
211             the commonality is held almost in the condition of slaves,
212             and dares to undertake nothing of itself, and is admitted
213             to no deliberation.</seg>
214         </p>
215     </div>
216 </div>
217 <div xml:id="div-french_1" xml:lang="fr">
218     <div type="section" n="6.13">
219         <head>XIII</head>
220         <p>
221             <seg xml:id="french_1">Partout en Gaule il y a deux
222             classes d'hommes qui comptent et qui sont considérés.
223             Quant aux gens du peuple, ils ne sont guère traités
224             autrement que des esclaves, ne pouvant se permettre
225             aucune initiative, n'étant consultés sur rien.</seg>
226         </p>
227     </div>
228 </div>
229     <linkGrp type="alignment" domains="#div_latin_1 #div-english_1
230                                     #div_french_1">
231         <link target="#latin_1 #english_1 #french_1"/>
232     </linkGrp>
233 </body>
234 </text>

```

As can be seen from lines 229-32 in listing IV.3, ekdosis uses the linkGrp mechanism as it is

Conspectus siglorum

Familia $\alpha$		
A	<i>Bongarsianus</i> 81	s. IX–X
	A <sup>1</sup> <i>Emendationes scribae ipsius</i>	
M	<i>Parisinus Lat.</i> 5056	s. XII
B	<i>Parisinus Lat.</i> 5763	s. IX–X
R	<i>Vaticanus Lat.</i> 3864	s. X
S	<i>Laurentianus</i> R 33	s. X
L	<i>Londinensis</i> Br. Mus. 10084	s. XI
N	<i>Neapolitanus</i> IV, c. 11	s. XII
Familia $\beta$		
T	<i>Parisinus Lat.</i> 5764	s. XI
f	<i>Vindobonensis</i> 95	s. XII
U	<i>Vaticanus Lat.</i> 3324	s. XI
l	<i>Laurentianus</i> Riccard. 541	s. XI–XII

1	XIII. In omni Gallia eorum hominum	XIII. Throughout all Gaul there are	XIII. Partout en Gaule il y a deux
2	qui aliquo sunt numero atque honore ge-	two orders of those men who are of any	classes d'hommes qui comptent et qui sont
3	nera sunt duo. Nam plebes paene seruo-	rank and dignity; for the commonality is	considérés. Quant aux gens du peuple, ils
4	rum habetur loco, quae nihil audet per se,	held almost in the condition of slaves, and	ne sont guère traités autrement que des
5	nullo adhibetur consilio.	dares to undertake nothing of itself, and	esclaves, ne pouvant se permettre aucune
		is admitted to no deliberation.	initiative, n'étant consultés sur rien.

<sup>2</sup> aliquo z] in al-  $\beta$  4 nihil audet (aut et A<sup>1</sup>) per se AM] nihil habet per se RSLN per se nihil audet  $\beta$  5 nullo z] nulli  $\beta$  consilio] conc- T U

Figure 4: Caesar, *BG*, VI, 13.1: Latin, English and French

described in the current TEI P5 Guidelines<sup>21</sup> to express alignment of parallel or translated texts. The three environments that have been set with `\SetEkdosisAlignment` (see listing IV.1) have been assigned automatically generated `xml:ids`, with the language attributes that were specified as further optional arguments as can be noted from listing IV.3, ll. 178, 205 and 217. Then, as the correspondences are between spans of text,  $\LaTeX$  `segment` environments have been translated into TEI `<seg>` elements each of which have been assigned again an automatically incremented `xml:id`. Finally, the segments found in these separate domains are connected together by means of `<linkGrp>` and `<link>` elements.

This technique naturally applies to alignment of parallel texts in multilingual corpora, but is also well suited for parallel redactions and borrowings which can be further annotated using the `\note` command which ekdosis translates into TEI `<note>` elements with associated `type` attributes.<sup>22</sup>

## V CONCLUDING REMARKS

Fig. 4 shows Caesar's *Gallic War*, VI, 13.1 critical edition in print as it is typeset by ekdosis from the  $\LaTeX$  source file that has been commented on here. Arguably, this is how it should be read in a “non-actionable” form. Of course, displaying in like manner more texts or more translations would soon become impossible, not to say irrelevant. Yet ekdosis can select a handful of versions out of many and display them properly in print while building a database meant to stand for queries and extraction of data. Such is the spirit in which it was written.

## References

- Caesar. *Gallic War*, volume 2 of *Collection des Universités de France*. Les Belles Lettres.  
Digital Latin Library. Textual criticism, nodate. URL <https://digitallatin.org/library-digital-latin-texts/textual-criticism>.
- Jean Irigoin. *Tradition et critique des textes grecs*. Les Belles Lettres, 1997.
- Donald J. Mastronarde and Richard J. Tarrant. Review: Guidelines for encoding critical editions for the library of digital latin texts. URL <https://classicalstudies.org/scs-blog/donald-j-mastronarde/review-guidelines-encoding-critical-editions-library-digital-latin>.
- L. D. Reynolds and N. G Wilson. *Scribes and Scholars*. Clarendon Press, 3 edition, 1991.

<sup>21</sup><https://tei-c.org/release/doc/tei-p5-doc/en/html/SA.html#SACSAL>.

<sup>22</sup>See above n. 19.