

**R. M. Cleminson**  
(University of Portsmouth,  
Central European University)

## CREATING ELECTRONIC CRITICAL EDITIONS

**Abstract:** The paper deals with the creation in electronic form of editions of texts provided with an apparatus indicating variants. The method is to encode the entire text as an XML document with a very simple structure, dividing the text into a linear, non-hierarchical series of segments, each segment consisting of a section of text plus the variants to that section with an indication of the witness(es) in which they are found. XSLT is used to number the portions of text and the corresponding variants and then to extrapolate the variants into an apparatus. The result is an XML document with a relatively simple structure. Additional mark-up may be added at this stage if required, for example further automatic transformation into a TEI-conformant document.

**Keywords:** critical edition, variant, collation, XML, XSLT, TEI

The kind of document which we shall be considering here is one consisting of a text accompanied by an apparatus indicating variants to that text. Properly speaking, not all such editions are critical in the strict sense of the word; however, the term “critical edition” is popularly used as a sort of *pars pro toto* to designate all such documents, and, for want of a better expression, it is in this loose sense that we shall be using it here. We shall also confine ourselves to the creation of electronic critical editions *ab ovo*, rather than questions of the encoding of existing editions which have been created by other means.

Also beyond the scope of this contribution are the various collation programmes that allow the comparison of two or more electronic texts. Though these are a valuable and powerful tool, it is a prerequisite for their use that all the texts to be processed should first be encoded, which may not be either practicable or desirable. By no means every manuscript is sufficiently important to be worth encoding in its entirety, and to do so purely for the purposes of collation would be a disproportionate expense of time and effort.<sup>1</sup> Nor are all manuscripts kept in places or conditions where such work can conveniently be carried out. The usual practice has always been to collect variants and to add these progressively to the edition. In the course of this process the base text itself may be modified (unless it has been decided to use the actual text of a particular witness as the base text), and it is certain that the study of each new manuscript has the potential to transform the researcher’s understanding of the text and its transmission.

It is this traditional process that is to be automated. It is axiomatic (or ought to be) that the function of the new technology is to make the work easier, quicker and

---

<sup>1</sup> Cf. the comment in Bakker, p.30: “If the text tradition is largely stable and if one is not interested in analysing scribal habits, it could be more economical to collate in the traditional way.”

more reliable, and not to make it more complicated. We should therefore analyse the process of the traditional creation of a critical edition in order to see how the computer can help us.

One begins with a number of witnesses to a text.<sup>2</sup>

Вьшка нѣкаа въ тѣло нѣкогого велмѣжа въ мало врѣме крѣшесе, питѣясе крѣвѣвѣи его. И тихо плъзаѣщи невѣдома бѣше. Въ единѣ же ѣ ношеи, прѣиде гость еѣ бльха, ѣже напрасно и безъ ѣма ѣѣзвѣши спящаго мѣжа и пробѣди его.

Вьшѣка нѣкаа въ тѣлѣ нѣкогого велѣмѣжа ѣ въ малѣ врѣме криѣшесе, питающисе крѣвию его. И тихо плъзаѣщѣи невѣдомаа бѣ. Въ единѣ ношь прѣиде гостиѣ еѣ бльха, ѣже напрасно безъ разѣма ѣѣзвивѣшѣи мѣжа спеша ѣ пробѣди его.

Вошка нѣкаа ѣ нѣкогого велможи в тѣле в мало время крѣшесея питающиса крови его. И тихо ползаѣщи невѣдомаа бѣше. Въ единѣ же ѣ нощи приде гостиѣ еѣ блоха ѣже напрасно и безъ разѣма ѣѣзви спяща мѣжа и пробѣди его.

These can be collated to show where they agree and where there is variation, essentially a process of segmentation.

Вьшка нѣкаа | въ тѣло нѣкогого велмѣжа | въ мало врѣме крѣшесе, | питѣясе |  
 Вьшка нѣкаа | въ тѣлѣ нѣкогого велѣмѣжа | въ малѣ врѣме криѣшесе | питающисе |  
 Вошка нѣкаа | ѣ нѣкогого велможи в тѣле | в мало время крѣшесея | питающиса |

крѣвѣвѣи | его. И тихо плъзаѣщи невѣдома | бѣше | Въ единѣ | же ѣ ношеи |  
 крѣвию | его. И тихо плъзаѣщѣи невѣдомаа | бѣ | Въ единѣ | ношь |  
 крови | его. И тихо ползаѣщи невѣдома | бѣше | Въ единѣ | же ѣ нощи |

прѣиде | гость | еѣ бльха, ѣже напрасно | и | безъ | ѣма | ѣѣзвѣши |  
 прѣиде | гостиѣ | еѣ бльха, ѣже напрасно | безъ | разѣма | ѣѣзвивѣшѣи |  
 приде | гостиѣ | еѣ блоха ѣже напрасно | и | безъ | разѣма | ѣѣзви |

спящаго мѣжа | и | пробѣди его.  
 мѣжа спеша | пробѣди его.  
 спяща мѣжа | и | пробѣди его.

(We are assuming that for our current purposes orthographic variation is irrelevant.)  
 On the basis of this we produce our critical edition:

Вьшка нѣкаа<sup>1</sup> въ тѣлѣ нѣкогого велмѣжа<sup>1</sup> въ мало врѣме крѣшесе, питающиса<sup>2</sup> крѣвию<sup>3</sup>  
 его. И тихо плъзаѣщи невѣдома бѣше<sup>4</sup> Въ единѣ же<sup>5</sup> ѣ ношеи<sup>5</sup> прѣиде гостиѣ<sup>6</sup> еѣ бльха,  
 ѣже напрасно и<sup>7</sup> безъ разѣма<sup>8</sup> ѣѣзвивѣшѣи<sup>9 10</sup> спящаго мѣжа<sup>10</sup> пробѣди его.

1-1 ѣ нѣкогого велможи в тѣле С  
 2 питѣясе Ка  
 3 крѣвѣвѣи Ка крови С  
 4 бѣ Л  
 5-5 ношь Л  
 6 гость Ка  
 7 ом. Л

<sup>2</sup> In this particular case, the manuscripts are, respectively, Belgrade, Patriarchal Library, MS 163, Belgrade, University Library, Lesnovo Monastery (Ćorović) collection, MS 31, and Moscow, Historical Museum, Synodal collection, MS 367. For the sake of clarity diacritical marks have been omitted in transcription.

8 □ ма Ка  
 9 □ □ зв □ ши Ка □ □ зви С  
 10-10 м □ жа спеща Λ add: и Ка С

A structure is thus imposed on the text which presents it as a series of segments, to each of which one or more variants corresponds. It also includes information indicating the origin of each variant. The initial or base encoding of the text should be constructed in a manner reflecting this structure, thus:

```
<text>
<s><t>Вышка н□каа </t></s>
<s><rb/><s><t> въ т□ло н□кого велм□жа </t></s><re/><v><r w="C"> □ н□кого
велможи в т□ле </r></v></s>
<s><t> въ мало вр□ме крїашесе, </t></s>
<s><t> питающихся </t><rp/><v><r w="Ka"> пит□ясе </r></v></s>
<s><t>кръвию </t><rp/><v><r w="Ka"> кръвї□вї □ </r><r w="C"> крови </r></v></s>
<s><t> его. И тихо плъза□щи нев□дома </t></s>
<s><t> б□ше </t><rp/><v><r w="Λ"> б□ </r></v></s>
<s><t> Въ един□ </t></s>
<s><rb/><s><t> же □ ноцей</t></s><re/><v><r w="Λ"> ноць </r></v></s>
<s><t>прїиде </t></s>
<s><t> гости□ </t><rp/><v><r w="Ka"> гостъ </r></v></s>
<s><t> е□ бльха, □ же напрасно </t></s><s><t> и </t><rp/><v><r
w="Λ">om.</r></v></s>
<s><t> безъ </t></s>
<s><t> раз□ма </t><rp/><v><r w="Ka"> □ма </r></v></s>
<s><t> □ □ звив □ ши </t><rp/><v><r w="Ka"> □ □ зв □ ши </r><r w="C"> □ □ зви
</r></v></s>
<s><rb/><s><t>спящого м □ жа </t></s><re/><v><r w="Λ"> м □ жа спеща </r><r w="Ka
C">add: и </r></v></s>
<s><t> проб□ди его. </t></s>
<witList>
<witness id="Ka">Belgrade, Patriarchal Library, MS 163</witness>
<witness id="Λ">Belgrade, University Library, Lesn.31</witness>
<witness id="C">Moscow, Historical Museum, Syn.367</witness>
</witList>
</text>
```

Each segment <s> consists of a portion of the text <t> together with any variants to that portion of text. If the text of a segment to which there are variants consists of a single word, it is followed by an <rp> tag, and then the variants; if it is longer, so that both its beginning and its end will have to be flagged in the edition (like the first, fifth and tenth variants in the example above), it is preceded by <rb> and followed by <re>. The <v> element includes the variants, each reading enclosed within an <r> element with a mandatory w attribute indicating its source. Though the encoding is minimal (only ten elements are declared), it is sufficient to generate a critical edition: in fact the sample critical edition above was generated automatically, without further manual intervention, from this very document instance by means of XSLT.

The transformation from primary encoding to critical edition has three stages. In the first, the variants are numbered, by means of consecutively numbered attribute-values attached to the <rp> and <rb> tags, and all the <rp>, <rb> and <re> elements are given unique location identifiers. In the second, the number of each <rb> element is assigned to its corresponding <re> elements, and the numbers and location identifiers of the <rp>, <rb> and <re> elements are assigned to the corresponding <v> elements. In

the third stage the <v> elements are removed from the body of the text and gathered together to form the apparatus. All three stages can be accomplished in a single operation by combining the three commands in a batch file.

The advantages of this method are particularly evident when changes are to be made to the edition. Supposing that another manuscript becomes available and its variants are to be added, we find ourselves dealing with this text:<sup>3</sup>

Гл□□т□босе □ко нека ваш□ка въ т□ле неког□ вел□м□жа въ мало вр□ме крїашесе,  
питающисе кръвію его. И тихо плъзаючи н□в□дома б□. Въ един□ же ноць прїиде къ н□и  
гостї□ е□ бльха, □же напрасно без□ раз□ма □□звив□ши м□жа сп□ща проб□ди его.

It is immediately obvious that the first variant occurs at the very beginning, which means that every subsequent variant will have to be renumbered. If this were to be done manually it would be an irksome and laborious task—over a long text prohibitively so—and also one very prone to introduce errors. If, however, the variants are added to the primary encoding, so that it begins

```
<s><t><rp/><v><r w="Ак">add: Гл□□т□босе □ко </r></v></s>
<s><rb/><s><t>Въшка н□каа </t></s><re/><v><r w="Ак">нека ваш□ка</r></v></s>
<s><rb/><s><t> въ т□ло н□кого вел□м□жа </t></s><re/><v><r w="С"> □ н□кого
вел□можи в т□ле </r></v></s>
```

and the transformation repeated, the variants will be renumbered automatically, without any danger of the linkage between text and apparatus being disrupted, as can so easily happen when manual additions and corrections are made. This means that the edition can always be revised, if errors are noted or additional information becomes available, and the very considerable labour and risk of introduction of errors entailed by this process is eliminated. It also eliminates the need to wait until all variants from all witnesses are collected before constructing an edition: the possibility of generating one at any stage of the work is an extremely valuable resource for anyone who is studying the history of a text and the relationships between its manuscripts. (“Lachmann’s Circle”: the significance of a variant is evaluated by reference to the established text, which itself is established through the evaluation of variants; both are subject to continuous reassessment in the light of the accumulation of information.)

The process could end here; or additional mark-up could be added if required. In the critical edition, as in the base encoding, the position of variants within the text is indicated by empty elements (<rp/>, <rb/>, <re/>) – in other words there is no *structural* mark-up. This avoids the danger of a conflict of structures which might otherwise arise if some sort of hierarchical division of the text is introduced at this stage. It is for this reason that the mark-up of the primary encoding is limited to the minimum required to indicate textual variation. Consider the text of Heb. vi 13-14:

Авраам□ бо □б□това б□ъ, понеже ни □дин□мъ большимъ им□аше кляти се. клятъ ся  
собо□ г□ля въистин□. блс□гвя бл с□гв□ тя и мнозя □множ□ тя.

авраам□ □б□това б□ъ, понеже большимъ един□мъ им□аше клетисе, клет бо се г□ле съ  
соб□ю въистин□ бл с□ве бл с□веще□ те и □множ□ □множ□ те

In comparing variants from these two sources, one would certainly identify as a segment собо□ г□ля/г□ле съ собо□ю. However, собо□ is the last word of verse 13 and г□ля the first word of verse 14; consequently if the text were marked up structurally before

<sup>3</sup> From Belgrade, Serbian Academy of Sciences, MS 25.

segmentation, overlapping elements would result. Structural or hierarchical mark-up at this stage should therefore be avoided.

If the critical edition is required for local use, one may end the process here, but if some sort of interchange of texts is envisaged, then a standard is required. It is, again, extremely easy to convert the encoded critical edition into a TEI document—in fact it involves little more than the renaming of some of the elements—and this can be achieved by a single XSL transformation: all that needs to be done manually is the addition of the required parameter entity references in the prologue, since this is something that XSLT cannot do. The result is a legal if not entirely legitimate TEI document, since it includes a mixture of location-referenced and double-end-point-referenced variants, something permitted by the TEI DTD but not apparently envisaged by its authors.<sup>4</sup> If a more purist approach to the TEI is preferred, this may be achieved by inserting an additional stage at the beginning of the transformation of the primary encoding. This will convert all location references to double-end-point references, and at the same time add an attribute to the element marking the front of the segment (<rb/>) to indicate whether it consists of a single word or a longer span. The reason for differentiating between these two types of segment is that in the visual rendition of the critical edition an indication of the beginning of a longer span is necessary, but for a single word it is redundant.

In principle it would have been possible, of course, to have used the TEI from the outset, but this would have lost the advantage of simplicity in the primary encoding. The TEI was never intended as an authoring tool, and is far too complex to function well as one. In particular, for documents which are intended for further processing rather than presentation, one does not want to have to contend with issues of conformance to a larger scheme which are not related to one's immediate purposes. This is true whether one is designing the document structure or validating a document instance. It should also be borne in mind that one of the advantages of XML is that files are suitable for multiple use. The embedding of variants within a text may have several possible purposes, of which the production of a publicly available critical edition is only one. It could also, for example, be used for statistical purposes, or various types of quantitative codicology. All of these will have outputs in an appropriate format, which need not be the same for each. Meanwhile the primary encoding remains in a format reflecting the four basic principles of multiple use, structure, portability and preservation proclaimed in the early days of computer-assisted processing of early Slavonic texts.<sup>5</sup> In the ten years since these principles were enunciated, the possibilities for the first, multiple use, have been greatly enhanced by the development of XSLT, which allows files to be

---

<sup>4</sup> See the discussion of linkage in Sperberg–McQueen and Burnard, ¶19.2. (TEI P5, which is still under development as this paper is being written, does not appear to be introducing significant changes in this area.) It is worth considering that although location-referenced linking is convenient for indicating variants to single words (it simplifies the primary encoding and the final stylesheet), the TEI does not envisage its use for the type of apparatus that we have been discussing hitherto, but rather for the type with which we are familiar from our copies of the Greek New Testament, where the apparatus is divided in parallel to the divisions of the text and the variants grouped accordingly, but the exact position of each variant is not indicated in the body of the text. Though one could produce an apparatus of this type by means of the method just described, it would involve considerably more labour (given the need to incorporate some sort of reference system in the primary encoding) for considerably less advantage, since subsequent additions and changes to this type of apparatus have no repercussions elsewhere in the document.

<sup>5</sup> See Birnbaum *passim*.

converted automatically, without loss of data, for different purposes. The transformations described in this paper are a small example of its potential.

The xsl scripts and other files necessary to perform the transformations described in this article may be found [here](#).

### References

- [1] H. P. S. Bakker, *Towards a Critical Edition of the Old Slavic New Testament*, Amsterdam, 1996
- [2] David Birnbaum, "How Slavic Philologists should use Computers", *Компютърна обработка на средновековни славянски ръкописи: доклади. Първа международна конференция, 24–28 юли, 1995, Благоевград, България, София, 1995*, pp.20-28
- [3] C. M. Sperberg-McQueen and Lou Burnard, *TEI P4: Guidelines for Electronic Text Encoding and Interchange*, Oxford, &c., 2001

[ralph.cleminson@port.ac.uk](mailto:ralph.cleminson@port.ac.uk)