Преглед НЦД 5 (2004), 31–41

Tomaž Erjavec
(Jožef Stefan Institute, Ljubljana)
Matija Ogrin, Jože Faganel
(Institute of Slovenian Literature, Ljubljana)

E-SLOMŠEK: A TEI ENCODING OF A CRITICAL EDITION OF 19TH CENTURY SLOVENIAN RHETORIC PROSE

Abstract: We describe the first Slovenian electronic critical edition, "The Three Sermons on Language" by Anton Martin Slomšek (1800-1862), Slovenian bishop, reformer and man of letters. The e-edition comprises digital facsimiles of the manuscripts, diplomatic transcription, critical transcription and apparatus. The e-edition is based on XML technology and all of the components are marked-up following the Text Encoding Initiative Guidelines, TEI P4. It was published as a result of cooperation between two Slovenian research institutions, the Scientific Research Centre of the Slovenian Academy of Sciences and Arts and the Jožef Stefan Institute; it is freely available at http://nl.ijs.si/e-zrc/slomsek/.

Key words: textual criticism, critical editions, XML, Text Encoding Initiative, Anton Martin Slomšek, 19th century Slovenian literature

1. Introduction

In a small nation such as Slovenia, publishing critical editions with facsimile, transcriptions and apparatus in traditional print form faces great economic barriers, primarily due to the very small book market. In such circumstances critical editions in electronic medium have a much better chance of carrying on the mission of preserving, interpreting and making available Slovenian cultural heritage – and thus contributing to our national identity. Yet, for this purpose, an elaborated methodology is needed, which will, on the one hand, encompass the specific editorial problems of Slovenian literature and, on the other, be based on open international standards and guidelines for text encoding and interchange. This combination is the goal of the collaborative project between the Institute of Slovenian Literature and Literary Sciences of the Scientific Research Centre of Slovenian Academy of Sciences and Arts and the Department of Knowledge Technologies at the Jožef Stefan Institute in Ljubljana, Slovenia. In this paper we present the result of the first stage of our project, i.e. "e-Slomšek", a text-critical edition of three sermons by Anton Martin Slomšek.

During the time of Anton Martin Slomšek (1800–1862; bishop, reformer, pedagogue and man of letters), Slovenia was a part of the Austro-Hungarian Empire. Throughout the 19th century, the Slovenian language was threatened by Germanisation – in some areas it was even in danger of extinction. In this climate Slomšek began a campaign of culture and education: he founded new schools, encouraged Slovenian language and culture, wrote and edited textbooks, and

published his sermons and episcopal statements with theological argumentation for national individuality. Slomšek's cultural activity was of great extent; he contributed significantly to the growth of literacy and reading in the mother tongue among Slovenians. Some historians maintain that the preservation of North-East Slovenia from complete Germanisation is mostly thanks to Slomšek [3].

In 2001, the book "Three sermons on language" was published [2], which contains the facsimiles of three sermons by Slomšek (two handwritten, one printed), and their diplomatic (for the first two) and critical (for all three) transcriptions by Jože Faganel. The diplomatic transcriptions closely reproduce the typography, historical spelling and syntax of the original documents, as well as the author's notes, emphasis, corrections and even mistakes, while the critical transcription brings the spelling, syntax and lexis somewhat closer to the modern Slovenian – following some clearly defined criteria – and introduces editorial corrections as well as critical notes.

In the project, the digital original of the book (in Microsoft Word) was, in a series of steps, converted into a standardised, self-describing format, namely the XML version of the Text Encoding Initiative Guidelines P4 [4], and its presentation for the Web. The project has, besides its obvious goal of producing widely accessible, flexible and interchangeable digital editions, also the purpose of bringing digital encoding expertise to the Academy partner. This had an impact on the work-flow in producing e-Slomšek, as well as on the shape of the project documentation. We should note that while quite a number of previous projects in the area of digital encoding of manuscripts and text-critical editions were studied before undertaking our own, the main inspiration was the William Blake Archive [6].

The rest of the paper is structured as follows: in the next section we present the workflow of converting and annotating the digital source of the book into the final TEI/XML form; section 3 describes TEI and the annotation structure of e-Slomšek; section 4 discusses the Web presentation of the materials, while the last section summarises and presents plans for the future. Appendix 1 gives a screen-shot of the HTML display of one page of the materials; this can serve to illustrate the form of the facsimile and electronic edition. The complete materials are available on the Web, at http://nl.ijs.si/e-zrc/slomsek/.

2. Converting and annotating the electronic edition

The production of the digital edition was one of step-wise refinement, where a series of well-defined steps normalises the data and mark-up. We can distinguish the following stages in the conversion to the final TEI P4 encoding:

1. preparation of the materials in MS Word
2. conversion to XML
3. conversion to basic TEI
4. first manual normalisation
5. pilot edition
6. second manual normalisation
7. addition of facsimiles
8. addition of editorial notes and hyperlinks
9. public edition

Below we focus on the more important step in this process. The TEI structure of the materials is explained in the next section.

2.1 Getting the material into TEI. The first step involved collating the MS Word documents, which were the basis of the printed version, into one file, which then served as the digital source for the TEI edition. From Word, the documents first needed to be converted to XML. There exist many programs that are able to perform this conversion, most of them from commercial vendors. However, there is an open source project, OpenOffice, which attempts to emulate MS Office, and includes an editor that is able to import MS Word format, but saves files in XML, which are valid according to its own OpenOffice Document Type Definition (DTD). In short, we imported the MS Word file into OpenOffice and saved it into its native XML format.

Using a combination of information obtained from the XML elements and attributes (in particular the paragraph and span style), we converted this encoding into a basic TEI structure. This meant first defining the TEI DTD for our project, which is obtained by parameterising TEI P4 (further discussed in Section 3).

A TEI parameterisation still allows for considerable leeway in the choice of the particular elements to use. It also contains a large number of elements not currently needed for the material. This is why we proceeded with the development of the encoding conventions by first defining the DTD via the TEI Guidelines / parameterisation and then manually writing a minimal, "strict" DTD, which specialises the TEI one, and is suitable for use in an XML editor and for developmental validation. The final interchange version of the materials then goes back to the full TEI DTD.

Conversion into this strict document type was done with a Perl filter, which also removes (or, in some cases marks with <sic>) all mark-up that it doesn't interpret as TEI elements, for example tab stops, font shifts etc. Unfortunately, as Word is not a constraining editor (i.e. identical visual effect can be obtained through a variety of means), this step lost a significant amount of information. The Perl filter also assigned numbers to the lines of the text; as will be seen, the linkage between the two transcriptions is via text lines, so their exact identification was therefore quite important – the automatic numbering, like the other automatic conversion steps, only correctly annotated a certain percentage of lines.

The next stage was the manual correction of the mark-up to obtain a consistently and correctly encoded TEI version. The first task here was to choose an XML editor; we considered several options (Emacs, JEdit, XML Spy, Oxygen), and in the end settled on Oxygen (http://www.oxygenxml.com/), which offers a good compromise between cost and usability. As the last preparation step we also wrote a stylesheet in XSLT, which converted the TEI/XML into HTML and attempted to reproduce the look of the printed edition.

Armed with the editor, DTD and stylesheet we then proceeded to correct and annotate the automatically derived TEI version. A first pass through the material corrected the obvious errors and ensured consistent line numbering, while all complicated cases were flagged with <sic> for subsequent discussion. To resolve these cases, a series of meetings was held, and advice sought esp. on the tei-l mailing list; the solutions we adopted forced a number of corrections and additions to the original encoding model, but enabled us to come up with, hopefully, sensible and consistent coding practices. The second manual pass through the materials resolved these

outstanding cases and resulted in the normalised and corrected version of the text. Finally, we automatically linked the lines of the diplomatic and critical transcriptions, whereby we arrived at the pilot TEI version of the texts.

2.2 A note on characters. A problem that was esp. apparent in the diplomatic edition, and that accompanies any edition of historical texts, is character encoding. While the printed or Web presentation of standard Slovene non-ASCII characters (č, š and ž and their upper-case equivalents – the "gajica" alphabet) is no longer a problem, this does not hold for historical characters used by Slomšek before 1846 (the "bohoričica" alphabet). Only a few years ago displaying these characters in a platform independent manner would have been impossible, but with the widespread adoption of Unicode it is, while, as will be seen, not completely solved, at least much more manageable. There were all together three special characters used in the diplomatic editions, and we discuss each one in turn.

A straightforward solution exists for ſ, a letter in common use in Roman types until the 18th century, and pronounced as the modern-day 's'. In Unicode it is described as LATIN SMALL LETTER LONG S, has the code point 017F, and belongs to the block Latin Extended-A. So, the character has exactly the right meaning and, furthermore, fonts for this block are available on practically all platforms.

A somewhat less robust solution exists for the mark Slomšek used to denote hyphenated words. Unicode does not have a character with the appropriate description, however, the character described as LOW DOUBLE PRIME QUOTATION MARK, code point 301F (CJK Symbols and Punctuation), is visually very similar to Slomšek's "hyphen". We therefore decided to adopt this character; still, not all platforms tested support it, as it is taken from the East-Asian character repertoire (CJK stands for Chinese, Japanese, and Korean). Therefore this character is in the derived HTML rendering displayed as "=".

Problems also arose with the LATIN CAPITAL LETTER LONG S, which does *not* exist in Unicode, even though it was, in Slovenian literature and probably elsewhere, a standard adopted upper-case equivalent of ſ. The problems with rendering this character were already evident in the printed edition, where it was approximated with the digraph ',S'. After experimenting with a number of solutions that attempted to emulate the printed edition and be "semantically clean", we, in the end, again decided to hijack a character visually close to what we wanted, in this case ∫, with the Unicode code point 222B, INTEGRAL (Mathematical Operators). The character seems to be supported on most platforms.

While XML supports the definition of entities which uniquely mark a character (say &Slong; for LATIN CAPITAL LETTER LONG S) this solution has the disadvantage that the entities must be defined in the DTD, and this prevents the documents from being distributed as stand-alone, i.e. without the DTD (so called "well-formed XML documents"). This is why we chose to store the special characters directly in Unicode; for the working version of the materials we used UTF-8 encoding while the public version uses plain ASCII, by using XML character entities. So, for example, 'ž' with code point 017E is stored as 'ž'.

2.3 The pilot edition. With the body of the materials in shape, we still needed to write the TEI header, a required element of any valid TEI document. The TEI header

contains the meta-data, i.e. it describes the document, and, in dedicated elements, provides information on various aspects of the data. After writing the header, it was joined to the body, with which the pilot XML edition of the resource was finalised. At this stage we wanted to test-mount the edition on the Web, in order to receive feedback on the usefulness of the resource and further check for errors and inconsistencies.

For the pilot Web version we re-wrote the XSLT stylesheet so that it also produced a table of contents by sermon and transcription, and, crucially, presented the two transcriptions in a line-by-line parallelised form. While the rest of the stylesheet essentially re-creates the look of the printed version, this last feature is a useful extension, both for the user, and for the editor in checking for inconsistencies between the two transcriptions.

At this stage an extensive report on the production of the TEI version was also written, giving the detailed steps of the conversion, as well as some more interesting discussions on our encoding dilemmas. Finally, an entry point HTML page was written, which also included an index to the images of the facsimiles of the sermons – these were, for the pilot version, not yet linked with the transcriptions. The material, in XML and HTML, the documentation, the images and the entry point were then mounted on the Web on a non-public URL, and this edition became the basis for the final revision.

2.4 Producing the final version. The third reading of the resources – esp. the parallel version – revealed some errors that had crept in during processing and editing, but also exposed mistakes originating from the original printed version. These were corrected and, simultaneously, references to passages from the Bible, written in the margins by Slomšek (e.g. "Luke 3.5") were identified and, in the critical transcription, corrected where necessary and hyperlinked to the corresponding passages of the Slovenian Web edition of the Bible (http://www.biblija.net/). Furthermore, editorial notes by M. Ogrin, regarding the genesis of original texts, the meaning of some archaic expressions and similar text-critical problems were added as notes in dedicated sections.

At this stage the facsimile was also incorporated into the TEI version, and the stylesheet was upgraded to display the TEI header and facsimile. The documentation, header and entry point were updated, and the final version of the resources mounted on the Web on the public URL.

3. The TEI structure of the material

E-Slomšek is encoded in TEI P4, the most comprehensive and widely used schema for text encoding. TEI P4 consists of the Guidelines, i.e. a prose description and the formal part, a set of XML DTD fragments (modules) that can be combined to arrive at a specific DTD suitable for a particular project. For e-Slomšek we used the following TEI modules:

- TEI.prose, the base module for prose, which inherits the TEI core consisting of the TEI header and elements for basic document structure, e.g. divisions, paragraphs, notes, emphasis, etc.

- TEI.transcr, an additional module for transcribing primary sources, which defines elements for correction that are used to encode the author's or editor's changes to the text.
- TEI.linking, an additional module, which we use to connect the facsimile and transcriptions.
- TEI.figures, an additional module, which we use to encode the facsimile.
- TEI.extensions, an optional module that contains project particular additions or changes to TEI. For e-Slomšek we make a few minor changes, e.g. add the attribute 'url' to various elements and constrain the 'rend' (rendering) attribute to a fixed set of values. More importantly, we also introduce two new elements, <page> and <line>, which will be further explained below.

With these choices we obtained our TEI DTD, and, while the parameterisation can be specified directly in the internal DTD subset of the document, we also made, with the help of the on-line TEI DTD generator, the so called "TEI Pizza Chef", a one-file DTD which is included with the distribution. As has been mentioned, we also created a strict DTD in the development phase, which specialises the official TEI one.

```
<profileDesc>
  <langUsage>
     <language id="en">angleščina</language>
     <language id="sl">slovenščina</language>
   </langUsage>
  <handList>
     <hand id="AMS" scribe="Anton Martin Slomšek"
       first="yes" />
     <hand id="JFA" scribe="Jože Faganel" />
     <hand id="MOG" scribe="Matija Ogrin" />
     <hand id="TER" scribe="Tomaž Erjavec" />
   </handList>
 </profileDesc>
```

Figure 1. An example from the TEI header

The complete structure of the materials is too complex to describe here in detail (but is explained in the on-line project report and in the TEI header of e-Slomšek) so we will mention below only some of the more interesting aspects of the encoding.

3.1 The TEI header. The purpose of the TEI header is to describe an encoded work so that the text itself, its source, its encoding, and its revisions are all thoroughly documented. Together these descriptions and declarations provide an electronic analogue to the title page attached to a printed work. The TEI header, tagged <teiHeader>, has four major parts:

1. a file description <fileDesc>, containing a full bibliographical description of the computer file itself; the file description also includes information about the source or sources from which the electronic document was derived;
2. an encoding description <encodingDesc>, which describes the relationship between an electronic text and its source or sources; it allows for detailed description of whether (or how) the text was normalized during transcription, what levels of encoding or analysis were applied, etc.;

3. a text profile <profileDesc>, containing classificatory and contextual information about the text, such as its subject matter, the individuals participating in producing it, and so forth. Such a text profile is of particular use in highly structured composite texts, where it is often desirable to enforce a controlled descriptive vocabulary;
4. a revision history <revisionDesc>, which allows the encoder to provide a history of changes made during the development of the electronic text.

To illustrate the kind of information contained in the e-Slomšek header, we give, in Figure 1 the profile description of the header, which defines the languages used in the header and text, and the hands used in the text – in TEI, a <hand> element is taken to identify each distinct person (author, editors) that contributed to the text. As will be seen below, we mark with the hand attribute corrections in the text, be it in the diplomatic or in the critical transcriptions.

```
<div id="sl1d" corresp="sl1k" n="1" type="dipl">
 <head>Diplomatični prepis</head>
  <page id="sl1d-f.1" corresp="sl1f.1" n="1">
   <line id="sl1d.1" corresp="sl1k.1" n="1"
     rend="right">1825. XIII</line>
   <line id="sl1d.2" corresp="sl1k.2" n="2"
     rend="center">Na 16 nedelo po Binkuſhtih.</line>
   <line id="sl1d.3" corresp="sl1k.3" n="3"
     rend="center">K’kerſhanſkimu govorjenju</line>
   <line id="sl1d.4" corresp="sl1k.4" n="4"
     rend="center">nagovor.</line>
   <line id="sl1d.5" corresp="sl1k.5" n="5">Takrat bode
tebi zheſt, kader tijiſti, kateri je tebe po&#301F;</line>
   <line id="sl1d.6" corresp="sl1k.6" n="6">vabil, tebi
porezhe: Prijatel, pomekni ſe gori.
    <note place="right">Luk. 14.</note>
   </line>
  ...
```

Figure 2. Begining of the diplomatic transcription of the first sermon

3.2 The text, divisions and pages. The text of e-Slomšek contains the <front> matter, comprising some introductory remarks, the <body>, and the <back> matter, with the bibliography referred to by editorial notes. The body contains three isions, one for each sermon. Each such first-level division then contains further s: one for the facsimile, one for the diplomatic and one for the critical transcription, as well as one containing editorial notes. The first two sermons have 4 divisions, while the last one, lacking the diplomatic transcription, has three. Each such division is marked-up with attributes for the type, number, correspondence and id. Each sermon also contains a “generated division” <divGen/>, which is an empty element, in essence a placeholder for the automatically derived division with parallel transcriptions.

A division consists of heads (not part of the sermons, but deriving from the printed version), and the individual pages. The structure here differs between the

facsimile (further explained later) and the transcriptions. For these latter, each page is encoded as a newly defined <page> element, composed of lines, <line>, also a local extension.[1] Both elements bear the correspondence and id attributes. This structure of the work is illustrated in Figure 2 by giving the first few lines of the diplomatic transcription of the first sermon.

We should note that this line-oriented encoding presents problems if the rhetorical (divisions, subdivisions) and linguistic (words) structure of the sermons were also encoded, as we are faced with the problem of overlapping hierarchies [5, 7]. However, the alternative, which is to encode lines with the empty TEI line break

```
<line id="sl1d.156" corresp="sl1k.156" n="156"><del
hand="AMS">tudi</del> dobro to ſvoje, in per temu <del
hand="AMS">pos'abijo</del> <add hand="AMS">blishenju
ſpregledajo</add> ſvojo laſtno </line>
…
<line id="sl1k.18" corresp="sl1d.18" n="18">tok bo tebi čest
v pričo povabl<add hand="JFA">j</add>enih vseh… </line>
```

Figure 3. Example corrections

element, <lg/>, presents problems for line-oriented processing.

3.3 The structure of a line. The lines carry the basic structure of the transcriptions, as they are the element that contains the actual text of the sermon, as well as notes and corrections. As mentioned, they also serve as the anchors which link the diplomatic and critical transcriptions, c.f. Figure 2.

The sermons contain a fair number of marginal notes that either mark the number of the "section" inside a sermon or refer to the Bible, i.e. the book and passage from where a particular quote was taken. The notes are marked with a rend attribute which specifies their position on the page, and, for references in the critical transcription, also with type="bibl". The lines furthermore contain mark-up for e.g. emphasised text <emph> (where Slomšek underlined the text) and gaps <gap/>, where the original was illegible.

3.4 Corrections. The transcriptions of the sermons also encode corrections, which occur as additions or deletions in the transcriptions. Those made in the diplomatic transcriptions emulate the ones on the facsimile, i.e. were made by Slomšek himself, either by crossing out a part of the text or writing new text above the old one. The corrections in the critical transcription were made by the transcriber and correct mistakes from the original. We use the same elements for both, namely <add>, <del>, and <corr>, but distinguish them by the hand attribute, defined as in Figure 1.[2] Two

[1] In the previous release of the materials the local <page> and <line> were encoded as the TEI line group, <lg>, and line, <l> elements. However, these are in the TEI defined as meaning the (grouping of) lines of verse. Using them for encoding manuscript pages and lines constitutes tag abuse (L. Burnard, M. J. Driscoll, p.c.).

[2] This encoding is problematic, as it conflates two fundamentally different "hands", the first one by the author, and present on the manuscript, the second one(s) by the editors and present only in the

examples, first from the diplomatic transcription and the second from the critical one, are given in Figure 3.

3.5 The facsimile. The originals of the first two (handwritten) sermons (8 pages each) were photographed on 6x9 cm slides, digitised in 300 dpi resolution and stored in TIFF format, while the third (printed) sermon was digitised directly into JPEG format. For the e-edition all sermons were stored in JPEG, to make for easier transmission, where image dimensions are approx. 800 x 1400 and size around 300 kB. The images were also resized to approximately a third of this side, to be able to display them in parallel with the text.

In the TEI encoded materials, as illustrated in Figure 4, the facsimile for each sermon occupies its own division , containing a <list>, where each <item> collects pointers to the images of one manuscript page in the different dimensions. Links to the image files are encoded as the (non-standard) url attribute of <figure>, which also specifies the type of the image.

```
<list>
  <item corresp="sl3k-f.1" id="sl3f.1" n="1">
    <note place="inline">Faksimile, pridiga III, stran 1</note>
    <figure type="jpeg" url="img/sl3-01.jpg" />
    <figure type="thmb" url="img/sl3-01_t.jpg" />
  </item>
...
```

Figure 4. Encoding the facsimile

4. Presentation of the material. While the focus of the effort at this stage was in producing the materials in a standardised encoding, we, of course, had to be able to present the end result. As mentioned, this was achieved by an XSLT stylesheet that converts the data into HTML. In addition to formatting the data in an almost identical manner to the printed edition, the stylesheet also renders the TEI header (in Slovenian), writes out the table of contents, displays the scaled-down facsimile next to the transcriptions, generates sections with parallel diplomatic and critical transcriptions, and hyperlinks the full-size images to the scaled down ones. A page in this rendering is given in Appendix 1.

Given the "didactic" nature of the project, special care was taken to make the documentation clear and comprehensive. The complete documentation is itself written in TEI (TEI Lite), and rendered in HTML via the TEI XSLT stylesheets. The TEI Guidelines P4 are also provided as a local copy, and each mention of a particular element in the documentation is linked to its description in the Guidelines, via an XML entity library. The documentation is thus exhaustive, and could, along with the materials themselves, be distributed stand-alone on CD-ROM. Finally, the documentation also contains an archive of written discussions and commentary, which additionally clarify some of the choices made in the encoding.

---

electronic edition. Under this view, the author's corrections should be marked by <add> and <del>, while the editor's corrections (emendations) by <corr> and <sic> (M.J. Driscoll, p.c.).

The Web HTML presentation of the sermons is currently static, i.e. the XML source is converted to HTML off-line via a single style-sheet. In the future, we plan to experiment with an configurable interface (maybe using AxKit or Cocoon XML modules for the Apache HTTP server), where users will have opportunity to choose the display mode, e.g. whether the corrections and emendations are visible as such, or folded into the text, what dimensions to display the facsimile in, which sermon and transcription to view, etc.

5. Discussion and further work

This paper presented the initial result of a common project between the Academy and Institute partners, the production of e-Slomšek. The focus of the work was first in developing the methodology and annotation scheme to produce standardised digital encodings of text-critical editions of Slovenian literature, but of course also in coming up with an interesting and usable result.

Nevertheless, there are many possibilities of improving our first offering, e.g. encoding the rhetorical and linguistic structure of the text. This second one is especially interesting, as tokenisation (mark-up of words and punctuation) enables the implementation of web-based concordancing engine (as already implemented for our Slovene-English parallel corpus [1]) as well as the extraction of a parallel (diplomatic/critical) lexicon from the text.

We are currently adding new works to our "e-ZRC" library, http://nl.ijs.si/e-zrc/, and have so far produced two other pilot editions. One comprises twenty-five letters from the correspondence of baron Žiga Zois (mostly from Jernej Kopitar), both influential figures in the awakening of national and linguistic identity in the 19th century Slovenia. The edition contains the facsimile (in German), the diplomatic transcription (transliteration from Gothic to Latin script), the translation into Slovenian and editorial comments. Furthermore, all the personal names in the edition were collected in a glossary, giving us a handle to the treatment of onomastics. The other pilot edition is the collection of "Songs about Maya" (*Pesmi o Maji*) written in the 1940's by Alojz Gradnik, a prominent Slovenian poet. The edition contains the poems in numerous variants, e.g. in first publication, in collected words, in the proofs, handwritten or typed versions, etc. and editor's comments. The edition thus enables exploring the relationship between the variants, among themselves and with a canonical version. Our further plans for the library include "Škofja Loka Passion Play" (*Škofjeloški pasijon*), and the Freising manuscripts (*Brižinski spomeniki*), the first having several video recordings, while the second one an extremely complicated critical apparatus.

References

[1] Tomaž Erjavec*: The IJS-ELAN Slovene-English Parallel Corpus*. International Journal of Corpus Linguistics, 7(1), pp.1–20, 2002.

[2] Jože Faganel (ed.) Anton Martin Slomšek: *Tri pridige o jeziku*. Mohorjeva družba, Celje, 2001.

[3] Jože Pogačnik: *Kulturni pomen Slomškovega dela*. Obzorja, Maribor, 1991.

[4] C. M. Sperberg-McQueen, Lou Burnard (eds.) *Text Encoding Initiative: Guidelines for Electronic Text Encoding and Interchange*, TEI P4, XML-compatible edition. TEI Consortium, 2002.

[5] Henry S. Thompson, David McKelvie: *Hyperlink Semantics for Standoff Mark-up of Read-only Documents*. In: Proceedings of SGML Europe '97, Barcelona, Spain, 1997.

[6] Joseph Viscomi: *Digital Facsimiles: Reading the William Blake Archive*. Computers and the Humanities, 36, p. 27-48, 2002.

[7] Patrick Durusau, Matthew Brook O'Donnell: *Implementing Concurrent Markup in XML*. Extreme Markup Languages, Conference Proceedings, Montréal, 2001.

tomaz.erjavec@ijs.si, matija.ogrin@zrc-sazu.si

Appendix I. The HTML display of the material