Philology

Digitalization of Georgian Medieval Manuscripts: Typicon of the Georgian Monastery of the Holy Cross Near Jerusalem

Irina Lobzhanidze

Institute of Linguistic Research, Ilia State University, Tbilisi, Georgia

(Presented by Academy Member Avtandil Arabuli)

The main reason for this paper is to underline significance of digital representation of Georgian medieval manuscripts in a form of online collections and to focus on specific issues regarding the technical approaches to digitalization process, role of metadata, delivery of content and facsimiles and worldwide access to the materials. The focal point of the paper is to represent digitization process taking for instance manuscripts represented in the Wardrops' Collection Online (WCO). The collection was prepared on the basis of Georgian manuscripts preserved in the Bodleian Libraries, University of Oxford. Georgian materials represented in the library belong to the period starting from the 8th up to the 19th centuries and can be subdivided into books and manuscripts represented in the Wardrop Collection and manuscripts from other collections (e.g. Lives of the Saints copied in Jerusalem in 1038-1040 etc.). For the purpose of this paper the Typicon of the Georgian monastery of the Holy Cross near Jerusalem, a manuscript of the 14th century, bought from the Church Missionary Society in 1911 by the Wardrop Fund is discussed. The work has been implemented on the basis of the Text Encoding Initiative (TEI) adopted for the representation of Georgian Medieval texts. The collection as well as the manuscript are freely available online and the approach discussed in the paper facilitates access to Georgian cultural heritage represented in a machine-readable format and allows us to cover complex and research-oriented aspects of Georgian medieval texts including stemma identification issues and preparation of critical editions online. © *2020 Bull. Georg. Natl. Acad. Sci.*

Digital edition, medieval manuscripts, editorial studies

The Bodleian library is one of the major repositories of Georgian Material abroad. Georgian materials represented by 210 items including 75 hand-written books belong to the period starting from the 8th up to the 19th centuries and includes a lot of rare and unpublished materials subdivided into those belonging to the Wardrops' collection and those from so-called Old Collection (e.g. Lives of the Saints copied in Jerusalem in 1038-1040 etc.). One of these manuscripts, Typicon of the Georgian Monastery of the Holy Cross near Jerusalem (MS. Georg. d. 2) was not published till 2019, but it was precisely described by Peradze [3] and then by Barett [4]. The Digital Bodleian Georgian collection includes photo material for six items. The Typicon of the Georgian Monastery of

the Holy Cross near Jerusalem is the last item added to the Digital Bodleian in 2019, and it became possible by common great efforts of Gillian Evison and Emma Stanford. In addition to image processing, data scanning and storage issues, the digitalization process of Georgian medieval manuscripts covers different stages including: 1) production of a diplomatic transcription, physical description and supporting commentary for each document in a form of meta-annotation scheme; 2) equipment of .xml files with critical apparatus and special tags indicating persons, places; 3) linking of manuscript images to the full text transcriptions; 4) preparation of .xml files for the further publishing to the web etc.

The Typicon is a good case study from different perspectives. Firstly, it is handwritten document, written in Asomtavruli and Nuskhuri scripts with Mkhedruli additions and some marginalia in the 14th century. It has not been published before neither in printed nor in digital forms. The Digital Bodleian providing online access to its photomaterial, but has not developed schemas for the transcription of texts.

Digitalization Principles of Georgian Medieval Manuscripts

The publication of manuscripts on the web means that there should be distinguished editorial model to be used and followed by the author of publication. The digitalization of Old and Middle Georgian manuscripts is associated not only with the definition of publication, but with concrete issues defined by the type of publications.

Raw text is a digital representation of text without images represented by means of any alphanumeric characters in Unicode or ASCII. UTF8 (Unicode) is a standard for all three types of Georgian scripts as Nuskhuri (Range: 2D00-2D2F), Asomtavruli and Mkhedruli (Range: 10A0-10F); its use does not need installation of additional fonts and does not cause problems with regards to the character representation on the web. The problem occurs if a server interprets Unicode characters differently and shows Georgian characters, especially, Asomtavruli and Nuskhuri used with Titlo Diacritics in a form of square;

Reading edition is a publishing format that allows readers to read a text in the form prepared by a concrete scholar or editor. But it creates a false impression that Old or Middle Georgian texts follow the structure and characteristics of modern ones consisting of letters, words, punctuation marks etc. In the majority of cases Old Georgian manuscripts never followed the above-mentioned structure of representation; punctuation marks were not represented at all except of dot (.) and three dots (∴) used to display sometimes the end of the paragraph, sometimes the end of the word; the use of white spaces was fragmentary, somehow, chaotic etc.;

Critical edition is a result of comparison between the fragments of manuscripts revealing the original or most significant form of a text. Generally, there are two possibilities: a) to represent a text of a concrete manuscript with comments or, b) to prepare a mixed version of textual fragments reconstructing the best text of a manuscript. Both of these approaches are widely used in publishing of Georgian manuscripts in printed or online versions and comprise the majority of problems already described with regards to reading editions;

Diplomatic edition is a transcription of a concrete manuscript preserving as much as possible the original reading, punctuation, line divisions, marginalia etc. Semi-diplomatic and diplomatic editions of Georgian manuscripts available online, generally, are based on their critical printed editions and share problems similar to those described previously;

Facsimile edition is a text preserved in the form of photography without any digital transcription of texts. This type of editions is not very frequent in case of Georgian manuscripts; at the same time, Georgian collection of Digital Bodleian can be considered as facsimile edition.

Following Piarezzo [5] the principle remains the same; all the texts are distinguished by different levels of editorial intervention in the text transmitted by a primary source and a document can be defined by its main characteristics like physical dimensions which can be measured (length, weight, number of leaves etc.), the place of origin and, of course, the text itself written in a concrete language. All, these dimensions are important, because the majority of digital editions are bound to mark-up and it is crucial to decide what is worth to be represented in the meta- and inner-annotation blocks. The workflow of our digital collection is as follows: Transcription – > Encoding (TEI, xml) – > Visualization (xslt, css). Transcription and mark-up of texts represented in Wardrops' collection are carried out against primary sources and aligned against facsimiles (if available).

The encoding procedure is subdivided into two stages: 1) meta-annotation stage, which includes thorough description of edition, sources, encoding etc. in accordance with schema used by Bodleian Libraries for cataloguing issues and amended by critical apparatus, lists of named entities (geographical and proper names), and, 2) transcription of texts in accordance with TEI:P5 Manuscript Description [6] including the referencing of text data to critical apparatus and of named entities to Wikidata [7].

In Wardrops' Collection Online [8], for each Georgian manuscript a single xml file was created separately. Files were stored in the open source native XML database eXist-db, processed by TEI Publisher [9] and connected to a web-site.

Typicon of the Georgian Monastery of the Holy Cross

According to Barret the Typicon of the Georgian Monastery of the Holy Cross near Jerusalem is written in the 14th century on paper in Khutsuri ecclesiastic script. The manuscript was bought from the Church Missionary Society in 1911 by the Wardrop Fund, described by Peradze in Georgian Manuscripts in England [3] and then by Barret in Catalogue of the Wardrop Collection [4]. It consists of 220 folios of 23 X 16 cm without title and end pages. Since 2019, the text of manuscript was not published [10], but according to Kekelidze [11] its provenance allows to suppose that it belongs to new edition of St. Sabbas monastic rules established since 10th-11th centuries.

By that time, the Georgian monastic rules generally followed the Great Svynaxar practices established by Giorgi of Mtatsminda [12]. The Great Svynaxar belonged to the typicon of the Great Church (Hagia Sophia) from Constantinople. It was translated from Greek by Giorgi of Mtatsminda and was used in Georgia since the 11th century. At this moment the Korneli Kekelidze Georgian National Centre of Manuscripts possesses the Great Svynaxar preserved under the shelf mark H-2211 [13] and about 18 editions of Typicon. The well-known publication of the Typicon of Shio-Mgvime Monastery [14] is based on different manuscripts (A – H 1349 (1247-1249 years) etc.).

The publication of Kochlamazashvili and Giunashvili is one of the best reading editions of the Typicon of Shio-Mgvime Monastery. Its text is represented in Mkhedruli script and, strictly, follows the structure and characteristics of modern Georgian texts consisting of letters, words, punctuation marks and white spaces with clear understanding of chapters, paragraphs etc. making it readable to modern readers.

As Kochlamazashvili states, the work started by Elguja Giunashvili in 1990 on the basis of H-1349 and appended with fragments from 18 manuscripts by him in the form of critical edition. The fragments in published edition are not marked with critical apparatus; it means that it is impossible to understand the compilation strategies of text and to use it for identification of stemma. The edition encompasses some additional information on textual marginalia, but the format of their representation does not allow us to understand the

manuscript, the page, the place on page, the hand, the script of these marginalia. The lack of the above-mentioned information, does not allow us to conduct a comparative study of the text at the computational level.

Digitalization of Manuscript

A single xml file of the Typicon consists of the meta-data, a set of images and a full text of the manuscript: 1) <teiHeader> contains all metadata determined for a concrete manuscript; 2) <facsimile> contains a set of images; 3) <text> contains the text of manuscript.

The annotations schemas are thoroughly described in [8]. The Typicon includes information on 54 geographical places and on 451 saints thoroughly described by means of special tags. The facsimile (<facsimile>) contains a representation of the written source in the form of a set of images aligned towards the original text. The Typicon is linked to Digital Bodleian host through IIIF manifest [8].

The text of the Typicon reveals six different hands appearing mainly in marginalia. Each of these hands was described by <handDesc> element and pointed to by @hand attribute from within the text. Page breaks are defined in <pb> element, which has several attributes a @f attribute to represent folio number, a @n attribute to indicate the page number and a @facs attribute to link the image through IIIF manifesto.

To provide expanded versions of abbreviations met in the Typicon <am> element was used to show signs present in an abbreviation and <ex> element – to show letters added by an editor. Both these elements were used inside <choice> element to group a number of alternative encodings for the same point in a text. The Typicon encompasses 53348 expanded abbreviations.

The <sourceDesc> section was expanded by information on critical apparatus. The text was partially compared with the Typicon of Shio-Mgvime Monastery [14] to indicate sections different from those represented in the Typicon. But the best option is to compare the source text to the source manuscripts preserved in the Korneli Kekelidze Georgian National Centre of Manuscripts.

Using for instance the Typicon, it is outlined that the manuscript description module of Text Encoding Initiative is successfully adopted for Georgian Medieval manuscripts. At this moment the description of manuscript, its text written in Khutsuri and translated into Mkhedruli scripts aligned against the images of manuscripts is freely available online [8]. This approach facilitates access to Georgian cultural heritage represented in a machine-readable format and allows us to cover very complex and research-oriented aspects of Georgian medieval texts including stemma identification issues and preparation of well-structured critical editions online.

ფილოლოგია

შუა საუკუნეების ქართული ხელნაწერების დიგიტალიზაცია: იერუსალიმის ჯვრის მონასტრის ტიპიკონი

ი. ლობჟანიძე

ილიას სახელმწიფო უნივერსიტეტი, ლინგვისტურ კვლევათა ინსტიტუტი, თბილისი, საქართველო

(წარმოდგენილია აკადემიის წევრის ა. არაბულის მიერ)

აღნიშნული სტატიის მთავარ ამოცანას წარმოადგენს ქართული შუა საუკუნეების ხელნაწე-რების ციფრული დამუშავების შესაძლებლობების წარმოჩენა და იმ საკითხების განხილვა, რომლებიც ეხება დიგიტალიზაციის პროცესის ტექნიკურ მახასიათებლებს, მეტა-მონაცემების როლს, ძირითადი შინაარსისა და ფოტომასალის ინტერნეტში განთავსებასა და მასალაზე ხელმისაწვდომობის უზრუნველყოფას. სტატიის ძირითად მიზანს წარმოადგენს უორდროპების ონლაინ კოლექციის (WCO) დიგიტალური გამოცემის წარმოდგენა. კოლექცია მომზადდა ოქსფორდის უნივერსიტეტის ბოდლეს ბიბლიოთეკაში დაცული ქართული ხელნაწერების საფუძველზე. ბიბლიოთეკაში წარმოდგენილი ქართული მასალა მოიცავს პერიოდს მე-8 საუკუნიდან მე-19 საუკუნემდე და იყოფა ორ ნაწილად: წიგნებად და ხელნაწერებად, რომლებიც უორდროპების კოლექციის ნაწილია და სხვა კოლექციებში წარმოდგენილ ხელნა-წერებად (მაგ. 1038-1040 წლებით დათარიღებული იერუსალიმში გადაწერილი ჰაგიოგრა-ფიული კრებული და ა.შ.). სტატიაში განხილულია უორდროპების ფონდის მიერ ეკლესიის მისიონერული საზოგადოებისგან 1911 წელს შეძენილი მე-14 საუკუნის ქართული ხელნაწერი, ჯვრის მონასტრის ტიპიკონი. სამუშაო შესრულდა შუა საუკუნეების ქართული ხელნაწერების წარმოდგენის მიზნით ტექსტის კოდირების (TEI)-ს ჩარჩოს გამოყენებით. ინტერნეტში გან-თავსებული კოლექციაც და ხელნაწერიც ხელს უწყობს მანქანის მიერ წაკითხვად ფორმატში ქართული კულტურული მემკვიდრეობის წარმოდგენასა და შუა საუკუნეების ქართული ხელ-ნაწერების კომპლექსური და კვლევაზე ორიენტირებული მიდგომების განვითარებას, მათ შორის, სტემის დადგენასა და კარგად სტრუქტურირებული კრიტიკული ონლაინ გამოცემების მომზადებას.

REFERENCES

1. Peeters P. (1923) La prise de Jérusalem par les Perses, Beyrouth, Impr. catholique.
2. Garitte G. (1960) La prise de Jérusalem par les Perses, Louvain, Secrétariat du CorpusSCO.
3. Peradze G. (1935) Georgian manuscripts in England. *Georgica*, 1(1): 183-216, London, Georgian Historical Society.
4. Barrett D. (1973) Catalogue of the Wardrop Collection and of other Georgian books and manuscripts in the Bodleian Library, Oxford, Oxford University Press.
5. Piarezzo E. (2015) Digital scholarly editing: theories, models and methods, Burlington, VT: Ashgate.
6. Willcox P., et. al. (2019) TEI Consortium, eds. *TEI P5: Guidelines for Electronic Text Encoding and Interchange*. Version 3.6.0, retrieved from https://tei-c.org/guidelines/p5/.
7. Erxleben F., Günther M., Krötzsch M., Mendez J., and Vrandečić D. (2014) Introducing Wikidata to the Linked Data Web. *Proceedings of the 13th International Semantic Web Conference*, 50-65, New York, Springer-Verlag New York.
8. Lobzhanidze I. (2019) *Wardrops' Collection Online*. Version 1.0, retrieved from http://manuscript.iliauni.edu.ge/.
9. Meier W., Turska M. (2015) *TEI Publisher*. Version 5.0, retrieved from https://teipublisher.com/
10. Karanadze M., Kekelia V., Shatirishvili L., Ckhikvadze N. (2018) Georgian manuscripts abroad, Tbilisi (in Georgian).
11. Kekelidze K. (1980) History of Old Georgian Literature, Tbilisi (in Georgian).
12. Abuladze I. (1967) Life and Citizenship of Our Holy and Blessed Father, Giorgi of Mtatsminda. *Monuments of Old Georgian Hagiographic Literature,* 121-178, Tbilisi (in Georgian).
13. Dolakidze M., Chitunashvili D. (2018) Great Synaxarion, Tbilisi (in Georgian).
14. Kochlamazashvili E., Giunashvili E. (2005) Typicon of the Shio-Mgvime monastery, Tbilisi (in Georgian).