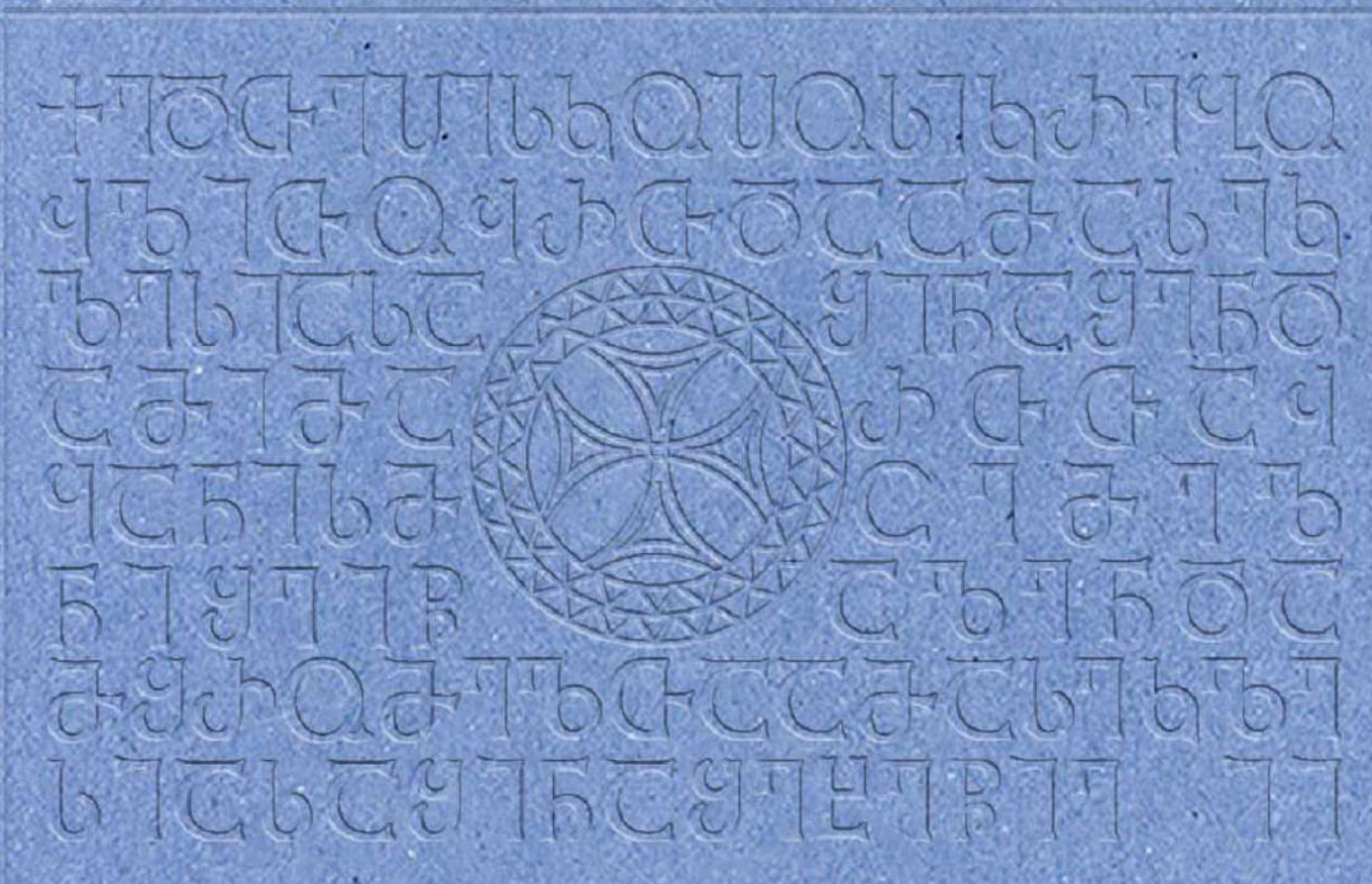


ACADEMY FOR DIGITAL  
HUMANITIES – GEORGIA

დიגיტალური  
ქართველოლოგია



**Digital  
Kartvelology**

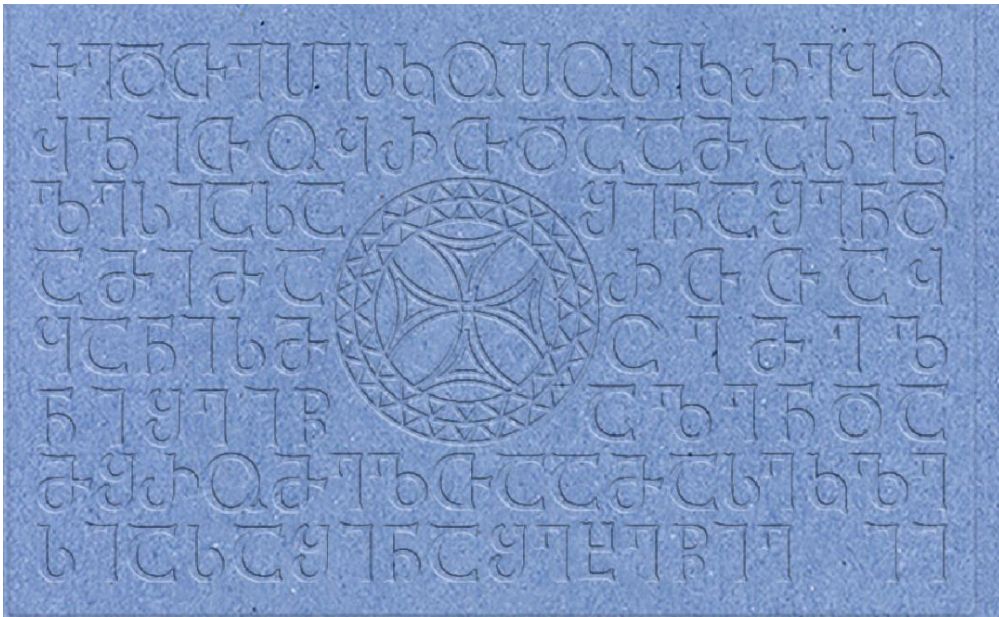
INTERNATIONAL BILINGUAL JOURNAL  
VOLUME 3



The Bilingual Scientific Online Journal of the **Academy for Digital Humanities**

დიגיტალური ჰუმანიტარის აკადემია - საქართველოს  
ორენოვანი საერთაშორისო სამეცნიერო ონლაინჟურნალი

# დიგიტალური ქართველოლოგია



## Digital Kartvelology

Volume 3

Dedicated to the Memory of Iza Chantladze (1.12.1939 – 6.6.2024)

ეძღვნება იზა ჩანტლადის (1.12.1939 – 6.6.2024) ხსოვნას

Tbilisi / თბილისი

2024

The journal **Digital Kartvelology** aims to promote the development of Digital Kartvelology and to create an academic platform for specialists working in Digital Humanities to publish scientific papers. The journal covers the following fields: Digital Humanities, Digital Kartvelology, Digital Caucasiology, Digital Lexicography, Translation Studies, Georgian Language Technologies, Corpus linguistics, Documentary Linguistics.

ჟურნალი **დიגיტალური ქართველოლოგია** მიზნად ისახავს დიგიტალური ქართველოლოგიის განვითარების ხელშეწყობას და წარმოადგენს აკადემიურ პლატფორმას აღნიშნული მიმართულების სპეციალისტებისთვის სამეცნიერო ნაშრომების გამოსაქვეყნებლად. ჟურნალში წარმოდგენილია შემდეგი დარგები: დიგიტალური ჰუმანიტარია, დიგიტალური ქართველოლოგია, დიგიტალური კავკასიოლოგია, დიგიტალური ლექსიკოგრაფია, თარგმანმცოდნეობა, ქართული ენის ტექნოლოგიები, კორპუსლინგვისტიკა, დოკუმენტოლოგია.

### Scientific Council:

Chairman of the Scientific Council: Jost Gippert (Germany)

Members of the Scientific Council: Manana Tandashvili (Germany), Paul Meurer (Norway), Bernard Outtier (Franch), Iryna Gurevych (Germany), Tony McEnery (England), Mzekala Shanidze (Georgia), Tinatin Margalitzadze (Georgia), Nino Pirskhalava (Georgia), Vakhtang Litcheli (Georgia)

Secretary of the Scientific Council: Mariam Kamarauli (Germany)

### სამეცნიერო საბჭო:

სამეცნიერო საბჭოს თავმჯდომარე: იოსტ გიპერტი (გერმანია)

სამეცნიერო საბჭოს წევრები: მანანა თანდაშვილი (გერმანია), პაულ მოირერი (ნორვეგია), ბერნარ უტიე (საფრანგეთი), ირინა გურევიჩი (გერმანია), ტონი მაკენერი (ინგლისი), მზექალა შანიძე (საქართველო), თინათინ მარგალიტაძე (საქართველო), ნინო ფირცხალავა (საქართველო), ვახტანგ ლიჩელი (საქართველო)

სამეცნიერო საბჭოს მდივანი: მარიამ ყამარაული (გერმანია)

### Editorial Board:

Manana Tandashvili (Editor-in-Chief), Darejan Tvaltvaдзе (Editor), Ramaz Khalvashi, Maia Lomia, Ketevan Margiani, Ketevan Datukishvili, Khatuna Beridze, Natia Dundua (Executive Secretary), Kakhaber Loria, Luka Nakhutsrishvili

სარედაქციო საბჭო: მანანა თანდაშვილი (მთავარი რედაქტორი), დარეჯან თვალთვაძე (რედაქტორი), რამაზ ხალვაში, მაია ლომია, ქეთევან მარგიანი, ხათუნა ბერიძე, ნათია დუნდუა (ადმინისტრაციული მდივანი), კახაბერ ლორია, ლუკა ნახუცრიშვილი



დიგიტალური  
ჰუმანიტარიის  
აკადემია

© Academy for Digital Humanities - Georgia

© დიგიტალური ჰუმანიტარიის აკადემია - საქართველო

ISSN: 2720-8427 (online)

# Digital Kartvelology, Vol. 3

## დიგიტალური ქართველოლოგია, ტომი 3

### Content/სარჩევი

#### I. Digital Manuscript Studies / ხელნაწერთა დიგიტალური შესწავლა

**Jost Gippert** (Hamburg), **Daniel Stökl Ben Ezra** (Paris) 5–24

Reading Georgian Manuscripts Automatically on the eScriptorium Platform

იოსტ გიპერტი (ჰამბურგი), დანიელ შტეკლ ბენ ეზრა (პარიზი)

ქართული ხელნაწერების ავტომატური ამოკითხვა eScriptorium-ის პლატფორმაზე

DOI: <https://doi.org/10.62235/dk.3.2024.8508>

**Maia Matchavariani** (Tbilisi) 25–40

Digital Humanities and the K. Kekelidze Georgian National Center of Manuscripts

მაია მაჭავარიანი (თბილისი)

დიგიტალური ჰუმანიტარია და კ. კეკელიძის სახელობის

საქართველოს ხელნაწერთა ეროვნული ცენტრი

DOI: <https://doi.org/10.62235/dk.3.2024.8509>

#### II. Digital Kartvelology / დიგიტალური ქართველოლოგია

**Hélène Gérardin** (Paris) 41–52

Building a Dialectological Lexicological Database of Georgian Cognates for Digital Analysis

ელენე ჟერარდინი (პარიზი)

დიალექტოლოგიური ლექსიკოლოგიური მონაცემთა ბაზის აგება

ქართული ენის ზიარი კოგნატების ანალიზისათვის

DOI: <https://doi.org/10.62235/dk.3.2024.8511>

**Ketevan Datukishvili, Nana Loladze, Merab Zakalashvili** (Tbilisi) 53–76

The Linguistic Model of the Georgian Language (Synthesis of Nominal Word forms)

ქეთევან დათუკიშვილი, ნანა ლოლაძე, მერაბ ზაკალაშვილი (თბილისი)

ქართული ენის ლინგვისტური მოდელი (სახელური სიტყვაფორმების

სინთეზი)

DOI: <https://doi.org/10.62235/dk.3.2024.8512>

#### III. Digital Rustvelology / დიგიტალური რუსთველოლოგია

**Manana Tandaschwili** (Frankfurt), **Mariam Kamarauli** (Hamburg) 77–108

Research into Equivalence in Multilingual Parallel Corpora

მანანა თანდაშვილი (ფრანკფურტი), მარიამ ყამარაული (ჰამბურგი)

ეკვივალენტობის კვლევისათვის მრავალენოვან პარალელურ კორპუსებში

DOI: <https://doi.org/10.62235/dk.3.2024.8514>



#### IV. Linguistics / ენათმეცნიერება

**George Hewitt** (London) 109–124

Paul's 1<sup>st</sup> Epistle to the Corinthians ch. 13. - A Comparison of Georgian and Abkhaz Translations (Taking into account the Greek Original)

ჯორჯ ჰიუიტი (ლონდონი)

პავლე მოციქულის პირველი ეპისტოლე კორინთელთა მიმართ - თავი 13 ქართული და აფხაზური თარგმანების შედარება (ბერძნული ორიგინალის გათვალისწინებით)

DOI: <https://doi.org/10.62235/dk.3.2024.8516>

**Maia Lomia, Ketevan Margiani** (Tbilisi) 125–136

The Theoretical Grounds and the Practical Value of Distinguishing the Markers of Evidentiality in the Verb Paradigms of the Kartvelian Languages

მაია ლომია, ქეთევან მარგიანი (თბილისი)

ქართველურ ენათა ზმნურ პარადიგმებში ევიდენციალობის მარკერის გამოყოფის თეორიული საფუძვლები და პრაქტიკული ღირებულება

DOI: <https://doi.org/10.62235/dk.3.2024.8517>

#### V. Lexicography / ლექსიკოგრაფია

**Tinatin Margalitadze, Salome Tchighladze** (Tbilisi) 137–150

Unknown Pages of English-Georgian Lexicography (George Ellis and his Comparative Lexicon of Kartvelian Languages)

თინათინ მარგალიტაძე, სალომე ჭიღლაძე (თბილისი)

ინგლისურ-ქართული ლექსიკოგრაფიის უცნობი ფურცლები (ჯორჯ ელისის ქართველური ენების შედარებითი ლექსიკონი)

DOI: <https://doi.org/10.62235/dk.3.2024.8518>

# Reading Georgian Manuscripts Automatically on the *eScriptorium* Platform

Jost Gippert (Hamburg), Daniel Stökl Ben Ezra (Paris)

DOI: <https://doi.org/10.62235/dk.3.2024.8508>  
[jost.gippert@uni-hamburg.de](mailto:jost.gippert@uni-hamburg.de) || ORCID: [0000-0002-2954-340X](https://orcid.org/0000-0002-2954-340X)  
[Daniel.Stoekl@ephe.psl.eu](mailto:Daniel.Stoekl@ephe.psl.eu) || ORCID: [0000-0001-5668-493X](https://orcid.org/0000-0001-5668-493X)

**Abstract:** The article outlines the development of means for an automatic reading of Georgian manuscripts on the *eScriptorium* platform and the first results achieved with them. After an overview of the efforts undertaken in applying Optical Character Recognition (OCR) to Georgian printed books since the late 1980's and a short introduction into the basics of the *eScriptorium* approach to Handwritten Text Recognition (HTR) and its functionalities, it exemplifies the application of the three core procedures of *eScriptorium*, which consist in the automatic segmentation of text-covered regions and lines, the automatic transcription of the detected lines based on manual input and the training of appropriate models, and the alignment with existing electronic texts in order to provide reliable ground truth for further training. With a total of 292 manually transcribed pages and 7488 pages with aligned (but not yet always corrected) text that have been processed so far, there is a strong material basis for further improvement of the models and the reading results depending on them.

**Keywords:** Optical Character Recognition, Handwritten Text Recognition, Text corpora, Old Georgian, manuscripts

## 1. From OCR to HTR

Nothing has changed the work of linguists, philologists and other researchers dealing with textual data as drastically as the availability of large corpora that can be searched and analysed digitally.<sup>1</sup> In the compilation of such corpora, two types of text data must be distinguished: those that exist in electronic form right from the beginning (i.e. that are “born digital”), and those that must be transferred into digital form (i.e. “digitised”) from other media, i.e. usually printed or handwritten sources.<sup>2</sup> The National Corpus of the Georgian language that has been compiled over the past 13 years (GNC, <http://gnc.gov.ge>) is a typical witness of this: the biggest of its subcorpora (the Georgian Reference Corpus [GRC], with 202,728,329 tokens) mostly consists of materials that were harvested in the World Wide Web, whereas the other subcorpora, especially those dealing with older stages of the language (GNC Old Georgian, with 7,101,021 tokens; GNC Middle Georgian, with 1,432,262 tokens; GNC Law texts, with 1,495,985 tokens), are more or less based on digitisations of printed matter that have been undertaken in the framework of the TITUS and ARMAZI projects.<sup>3</sup> Only in a few cases was the digitisation done manually, i.e. by typing a given text with the keyboard; in most cases the printed model was entered page by page via an optical scanner and then “read” into a digital text format via special software that was able to isolate the individual letters on the page, identify them with the

---

<sup>1</sup> On the general issue of text digitisation, see Stökl Ben Ezra (forthcoming). For an early programmatic approach concerning Georgian and other languages, see Gippert (1990).

<sup>2</sup> This includes text data from audiovisual media (e.g. recordings of spoken language) which must be transcribed before they can be integrated into corpora.

<sup>3</sup> See <https://titus.uni-frankfurt.de/texte/texte2.htm#georgant> and <https://armazi.fkidg1.uni-frankfurt.de>.



corresponding characters of the alphabet, and storing the sequences of letters as a coherent electronic text.

When the endeavours to compile a diachronic corpus of Georgian began in the late 1980s, this method, usually styled Optical Character Recognition (OCR), was by far not as well developed as it is today, and scanners were much less effective. The biggest problem consisted in the fact that personal computers of that time were not yet designed to deal with languages written in scripts other than Latin – they were first based on a 7-bit encoding system (ASCII) that comprised just the characters used in English, and this was only gradually substituted by an 8-bit system (ANSI) which also covered the “extra characters” necessary for “Western” languages like German (“umlaut characters”) or French (“accented characters”). With the introduction of the international standard named ISO/IEC-8859, Cyrillic and Greek became applicable alongside “North”, “Central” and “South European” Latin-based alphabets; however, the so-called “code pages” representing them did not imply a unique encoding, given that different characters had to be “mapped” across the set of 256 available code points (e.g., Greek  $\theta$  shared its code [232] with Cyrillic *u* and, depending on the code page, Latin  $\grave{e}$  and  $\grave{c}$ ).<sup>4</sup> For scripts like Georgian, there was no support at all, which means that in the attempt of dealing with Georgian texts, a mapping with Latin characters had to be devised at the beginning.

OCR, too, was strongly dependent on Latin at that time, which means that most software that came (or could be used) with the available scanners was preconditioned to recognise Latin characters, no others, and only those pertaining to the ASCII and, later, ANSI standards. It goes without saying that such software could not reasonably be applied for reading Georgian.<sup>5</sup> Even in 1988, however, there was already some OCR software available that could be “trained” to distinguish Georgian characters and to “map” them onto encodable Latin characters. One such software package which proved usable for this purpose and which was applied (in connection with a Xerox Datacopy scanner) for the digitisation of the first hundreds of pages of Georgian printed texts now stored in the GNC, was the “SPOT” program developed by a company named Flagstaff Engineering. Fig. 1 shows the functionality of the program, which was still based on the DOS operating system, in a set of screenshots including its “training” function.<sup>6</sup>

During the 1990s, another commercial product came on the market which provided a similar functionality but with higher efficiency; this was the “FineReader” program, now designed for Windows systems, which was developed by the Russian company ABBYY. Georgian did not belong to the many languages that the “FineReader” supported, nor was the Georgian script; however, the “mapping” principle could still be applied in both “training” and reading here, too, and so the bulk of the Old and Middle Georgian materials digitised in the course of the TITUS and ARMAZI projects were based on its application. It goes without saying that for the development of a consistent electronic corpus of Georgian, the “mapped” Latin-script renderings of Georgian characters (see Table I for an example) had to be corrected and then

<sup>4</sup> For a comprehensive survey see [https://en.wikipedia.org/wiki/ISO/IEC\\_8859](https://en.wikipedia.org/wiki/ISO/IEC_8859).

<sup>5</sup> This is true, e.g., for the widespread *OmniPage* software (first developed by Caere Corporation, see <https://en.wikipedia.org/wiki/OmniPage>) as well as the scanning systems of Makrolog (see <https://www.makrolog.de/home/>) and Kurzweil ([https://en.wikipedia.org/wiki/Ray\\_Kurzweil](https://en.wikipedia.org/wiki/Ray_Kurzweil)).

<sup>6</sup> Screenshots on the basis of the demo version in German.

converted into adequate codes; this became possible in the late 1990s with the development of the (32-bit) Unicode standard (ISO/IEC-10646), which meanwhile offers unique code points for all three Georgian scripts (*mkhedruli*, *nuska-khutsuri*, *asomtavruli/mrglovani*).

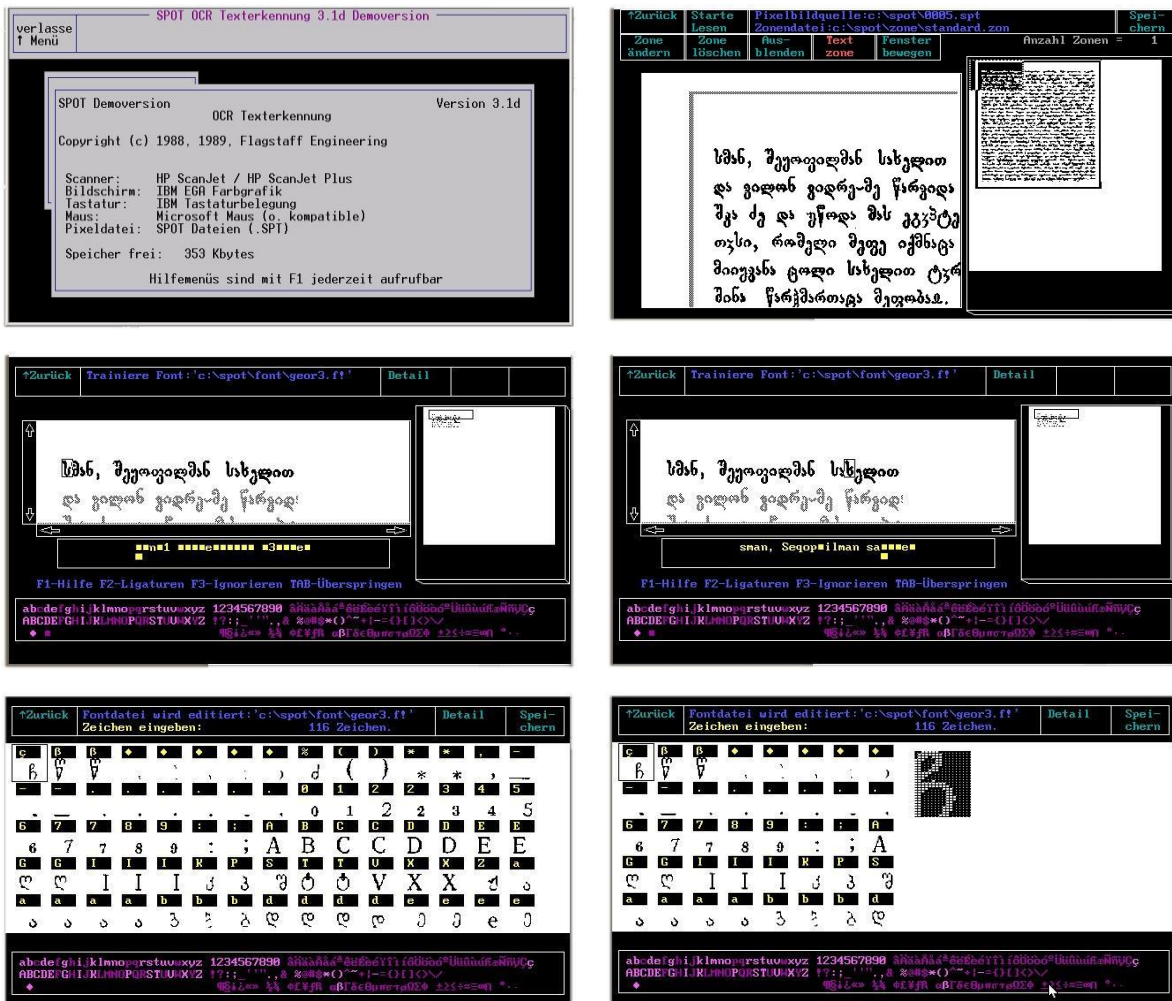


Fig. 1: DOS-based OCR Software (SPOT) with “training” function (screenshots)

Table I: OCR result with “mapping” onto Latin characters

's~ da ubr&ana didebul~a: amas Kacs~a saKv~irveli nadiri una-  
xavs, aC misi unaxaoba C~vengan ar egebis da, vinattgan esre-  
~aKv~irveli aris, mnebavs, ratam~a mteli SeviPQrato.  
gaiQara nadimi da SeeKazmes mepe da Qovelni didebulni  
mismi, Cina Cari&Gvanes igi Kaci da &ivides mas mi~-  
dorsa Sina. naxes igi kurciKi mit saxit~ masve adgilsa. br&a-  
na mepeman: aha male, aba, Qovelman Kacman cxenebi SeuTevet da  
siKvdilsa eKr&alenito. SeuTeves cxenebi da Cavida kurc~,Ki  
igi. arbives dGisa savali da mosCQda sxva Kaci Qvela, mepe da  
samni sxvani didebulni SerC~es da arbives vidre mCuxramd`,s  
~vidisa dGisa savali: verca mieCives da arca mosCQdes.  
adgilsa ertsa mivides Kldtvansa. C~aexveCa kurciki da  
uC~ino ikmna. odes daxednes kveQana~a, verasada cnes da ver~a  
naxes kalaki da verca sopeli, verca igi nadiri. kveQana  
ucxod auC~ndat; ara uCQodes, sadamca Cavides. mixed~ me-  
peman da naxa saxli erti kvitKirisa. gauKvirda da tkva: ra-



In this way, the Georgian National Corpus has incorporated the textual heritage of Old and Middle Georgian to a large extent, namely, as far as it has been available via printed editions. However, a great bulk of highly interesting materials are still awaiting their inclusion, namely, everything written in unpublished manuscripts. Attempts to automatically read text from images of manuscript pages with commercial OCR software yielded no acceptable results; the reason is that the variation of letter shapes is much higher in handwritten texts than in printed ones. For printed texts, it was necessary to undertake “training” for every single type face used in them; in the case of manuscripts, it would not even have been enough to undertake the same effort for every single scribe’s hand since the variation in character shapes is enormous even where scribes intended to be consistent in their handwriting style.

This problem has now been overcome by the development of AI-based trainable software for Handwritten Text Recognition (HTR), and the first endeavours to apply such a system to Georgian manuscripts undertaken in the framework of the DeLiCaTe project<sup>7</sup> during the past six months have yielded remarkable results. The system used is the *eScriptorium* platform, developed at the Paris Sciences et Lettres University as part of the projects *Scripta* and RESILIENCE with contributions from other institutions, partly funded by the EU’s Horizon 2020 funding program and a grant from the Andrew W. Mellon Foundation.

## 2. *eScriptorium*, its background and its functionality

*eScriptorium* is currently the only *cutting edge* open source HTR programme with an ergonomic interface.<sup>8</sup> It has been developed by the Ecole pratique des hautes études, Paris Sciences et Lettres University (PSL), in Paris since 2019 in collaboration with other teams, in particular the ALMAnaCH team of the Institut national de recherche en sciences et technologies du numérique (INRIA, National Institute for Research in Digital Science and Technology, Paris) and the openITI consortium of University of Maryland (College Park, MD), North-Eastern University (Boston/London), and Aga-Khan University (London).<sup>9</sup> Its central functionalities allow users to automatically analyse the layout of uploaded pictures of handwritten or printed documents and then to automatically transcribe them. This requires the existence of trained AI models for layout segmentation and text recognition of such documents, their script and/or language types. If there are no already trained adequate AI “models” available, users can manually create the necessary training material and train new models and/or fine-tune existing ones, even switching from one language or one script to another. The key to *eScriptorium*’s success is that this can be done relatively quickly and with less human effort than in the past.

With “models” we refer here to multilayered neural networks composed of convolutional neural networks (CNN) and bidirectional long-short-term memory neural networks (bi-LSTM). They are like mathematical formulae with millions of parameters that the computer optimises itself during the training process by being shown “ground truth”. For layout-segmentation, this

<sup>7</sup> *The Development of Literacy in the Caucasian Territories*, see 3.1 below.

<sup>8</sup> Its user-interface stack is available at <https://gitlab.com/scripta/escriptorium/>. The AI part is available at <https://github.com/mittagessen/kraken>. There are other open source HTR and OCR programs but their interface is by far not as ergonomic. And there are other HTR programs with an ergonomic interface but they are not open source or not cutting edge.

<sup>9</sup> See Kiessling et al. 2019: 19; Stokes et al. 2021: 18 (1).

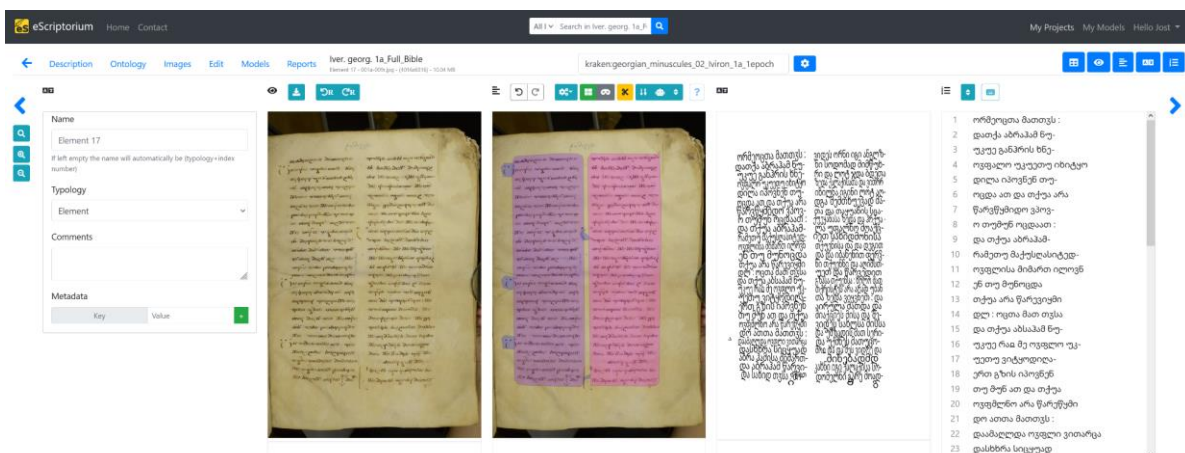
ground truth consists of document pages with annotations that describe the polygons of regions, line-contours and baselines as well as their types (e.g. main text, running header, etc). For recognition or transcription, the ground truth is Unicode-encoded strings of characters corresponding (and bound) to specific lines.

For the creation of ground truth and the correction of automatic layout segmentation or transcription, ergonomics are paramount. An unergonomic interface not only makes the correction process more time-consuming but also more tedious and more error-prone.

Data in *eScriptorium* is structured into projects which consist of documents, in their turn consisting of images (of pages, bifolios, or cutouts – as the user defines). Users can collaborate in teams and share the material among them. Each document image can have only one layout segmentation linked to it, but each segmentation can be linked with as many transcriptions as one would like, either automatically or manually, so that one can contrast e.g. several independent transcribers, keep abbreviations both unresolved and resolved, or have a normalised rendering alongside a diplomatic one.

Almost everything is user-definable. This means that users are free to define their own way to handle the layout segmentation and the typology of region- and line-types used. They can decide for their own transcription conventions, i.e. whether to expand abbreviations or leave them unresolved, whether to use a graphematic, a diplomatic or a normalising transcription and which conventions to follow, and which metadata to add on the document or image level. Transcriptions can be annotated and enriched with a versatile annotation interface where users can define what they want to annotate (e.g. named entities referring to places or people; dates; resolutions of abbreviations; connections between marginal or interlinear annotations and their insertion spots; etc).

In the edit mode, the user can currently choose to visualise one to five parallel panels: (1) metadata, (2) facsimile, (3) segmentation, (4) transcription and (5) text-annotation (see Fig. 2 for an example).<sup>10</sup> The facsimile (2), segmentation (3) and transcription (4) panels are twinned



**Fig. 2:** The five panels of *eScriptorium* in edit mode

<sup>10</sup> In the upcoming new interface created with the massive support of the openITI team, the metadata panel will become a modal window and the facsimile panel will be deprecated.



for zooming in/out or panning to visualise different document parts in the required enlargement. This enables the users to always see the specific region of the document image, its annotation and transcription in a what-you-see-is-what-you-get fashion. The segmentation panel serves the complex procedure of creating and correcting layout annotations, similar to a multilayer drawing programme (cf. Fig. 3).

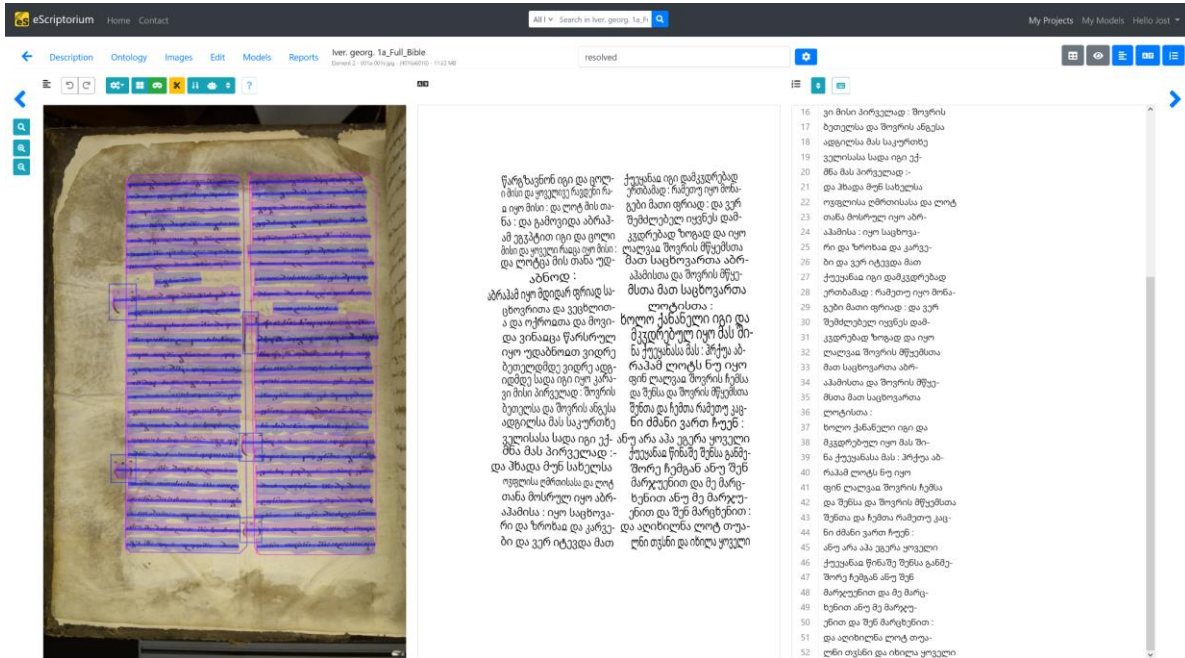


Fig. 3: Segmentation, transcription and text-annotation panels

In the transcription panel, clicking on a line opens it up in a modal window that displays the image of this line and its transcription directly beneath it so that the eye has the shortest distance between the image and the transcription (cf. Fig. 4). The text-annotation panel permits the users

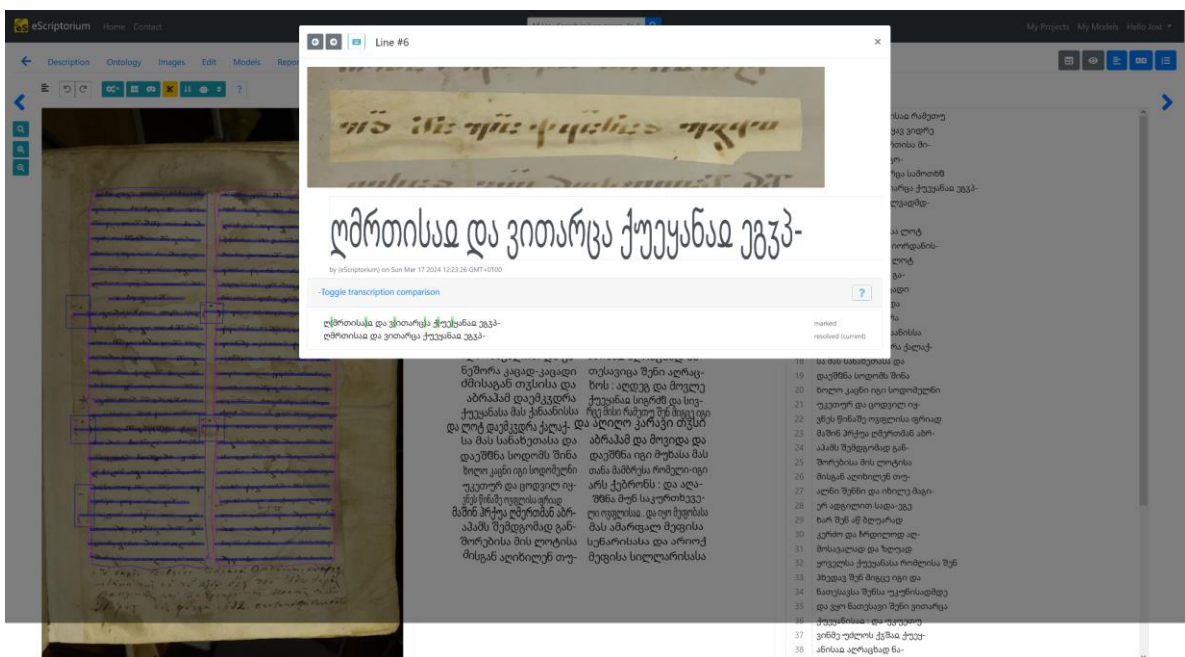


Fig. 4: Modal window for transcription

to create their own annotation buttons above it and then to add e.g. IDs for people, grammatical information, etc. It also allows users to correct the line-ordering.

A most important additional automatic feature is the text-to-text alignment based on the *passim* project by David A. Smith.<sup>11</sup> It allows e.g. to semi-automatically provide training ground truth from existing e-texts by first creating a rough automatic transcription and then aligning it with an existing gold standard transcription of a similar or the same text or manuscript.<sup>12</sup>

A powerful automated programming interface (API) permits to deal with almost all data points and functions automatically. Of course, images, annotations, and trained models can be exported, published and transferred between different instances of *eScriptorium*. The available formats for transcription export are (plain) text, Alto-XML<sup>13</sup> and Page-XML.<sup>14</sup> The latter two XML schemas include the layout coordinates and typology. In the near future, the text-annotation will also be directly exportable via METS.<sup>15</sup> Currently, one needs to use the API for this.

At present, excellent recognition models exist for languages written in Latin, Hebrew, Hindi, Chinese, Syriac, and Palaeo-Slavonic characters for certain periods. Various teams and projects are working on Greek, Avestan, Arabic, Japanese, Coptic, Armenian, Sanskrit, etc. In May 2021 we trained a first recognition model for printed Georgian (*mkhedruli*) on a small dataset consisting of 3668 lines. The model reached an accuracy of 99.2% on the test data.

*eScriptorium* can be installed on Linux and MAC-OS for individual users. A graphics processing unit (GPU) is needed if one wants to train layout-segmentation models. For the training of recognition models, a GPU is highly recommended as well. For teams, it is preferable to work on a more powerful server with a high speed internet connection. There are many configurations for different use cases and budget constraints, extending from small teams working on a single server with one GPU to very large multi-server high performance computing systems (HPCs) with multiple GPUs. If no GPU is available, we have also seen users exporting their data to a Google Colab(oratory),<sup>16</sup> installing *kraken* (the OCR system underlying *eScriptorium*)<sup>17</sup> and training the model there and then reimporting it. For large models created on data from multiple documents, we recommend exporting the data and training a model on an HPC cluster with a GPU because one has more control over hyper-parameters then. An active user-community on Gitter<sup>18</sup> helps new users and system-administrators to overcome eventual difficulties during the installation or initial use.

---

<sup>11</sup> See <https://github.com/dasmiq/passim>.

<sup>12</sup> Several parameters can be set for the alignment; for our Georgian manuscripts, the best results were achieved with a line length match threshold of 0.5, an N-gram value of 8 and a beam size of 100.

<sup>13</sup> “Analyzed Layout and Text Object”, see [https://en.wikipedia.org/wiki/Analyzed\\_Layout\\_and\\_Text\\_Object](https://en.wikipedia.org/wiki/Analyzed_Layout_and_Text_Object).

<sup>14</sup> “Page Analysis and Ground Truth Elements”, see [https://en.wikipedia.org/wiki/Page\\_Analysis\\_and\\_Ground\\_Truth\\_Elements](https://en.wikipedia.org/wiki/Page_Analysis_and_Ground_Truth_Elements).

<sup>15</sup> “Metadata Encoding and Transmission Standard”, see [https://en.wikipedia.org/wiki/Metadata\\_Encoding\\_and\\_Transmission\\_Standard](https://en.wikipedia.org/wiki/Metadata_Encoding_and_Transmission_Standard).

<sup>16</sup> See <https://colab.research.google.com/>.

<sup>17</sup> See <https://kraken.re/main/index.html>. We started using version 4.1.2 (see Kiessling 2022). The latest version used for this article is 5.2.9 (see Kiessling 2024).

<sup>18</sup> See <https://gitlab.com/scripta/escriptorium>.

In the following section, we will describe ongoing efforts undertaken within the DeLiCaTe project to use *eScriptorium* for the automatic reading of Georgian manuscripts from the first and second millennia of our era.

### 3. Using *eScriptorium* for Georgian Manuscripts: Achievements and Future Tasks

#### 3.1 The context

The DeLiCaTe project, an ERC Advanced Grant devoted to investigate the “Development of Literacy in the Caucasian Territories”,<sup>19</sup> has been running since April 2022 in the Centre for the Study of Manuscript Cultures (CSMC) of the University of Hamburg.<sup>20</sup> A decisive part of its objects consists of palimpsest manuscripts in Georgian, Armenian and Caucasian Albanian, which represent the oldest written sources available for the three languages of the Caucasus that have developed their own literacy in the course of the Christianisation of the region in the 4<sup>th</sup>–5<sup>th</sup> centuries CE. The decipherment of palimpsests, i.e. handwritten texts that were erased so that the writing support could be reused for writing down other text materials at a later time,<sup>21</sup> is usually extremely difficult and requires peculiar imaging methods, among which multispectral imaging (MSI) has meanwhile yielded the best results.<sup>22</sup> Even with these methods, however, many overwritten texts cannot be identified easily, especially when only a few words per page can be made out. Whenever the overwritten text is available from other manuscripts in electronic form and retrievable via search engines of corpora, a text passage can often be identified even if only a few characters are discernible; in the case of Georgian and Armenian, this is typically true of biblical matter.<sup>23</sup> With other relevant text genres such as hagiography, homiletics, and hymnography, the amount of digitally available text materials that can be used for searching is much more restricted, given that important parts of the written tradition of the Caucasus have never been edited and are only accessible in manuscript form to the present day. This is the reason why we decided in our project to apply *eScriptorium* in order to digitise and transcribe manuscripts that may contain relevant textual materials.

The manuscripts we are working on with *eScriptorium*<sup>24</sup> are usually parchment codices of the 9<sup>th</sup>–11<sup>th</sup> centuries of which colour images of sufficient quality are available, provided by (or procured from) the major repositories (for Georgian, the Korneli Kekelidze National Centre of Manuscripts, Tbilisi; the Iviron Monastery on Mount Athos; the University Libraries of Graz and Leipzig; the Bibliothèque nationale de France, Paris; the Bodleian Libraries, Oxford; and others). In some cases, we have to rely on greyscale images that were produced from microfilms; this is true, first of all, of the large collections kept in St Catherine’s monastery on

---

<sup>19</sup> European Research Council, grant agreement no. 101019006.

<sup>20</sup> See <https://www.csmc.uni-hamburg.de/delicate.html>.

<sup>21</sup> As to palimpsests in general and the DeLiCaTe approach, cf. Gippert (2025a and 2025b).

<sup>22</sup> Cf. Gippert (2025b), Kamarauli (2025), Kvirkevelia (2025), Sargsyan (2025) and Bonfiglio (2025) as to palimpsests that are investigated with multispectral imaging in the DeLiCaTe project and Mohammed, Jampour and Gippert (2025) as to a new method of reconstructing the erased layers of palimpsests.

<sup>23</sup> The decipherment of the Caucasian Albanian palimpsests, too, was only possible because the content could be identified as being biblical; see Gippert (2023: 99–141) for details.

<sup>24</sup> First attempts to use *eScriptorium* for reading the erased lower layers of palimpsests have failed – the system is optimised to ignore the traces of the overwritten text, treating them as “noise”.

Mount Sinai and the Greek Patriarchate in Jerusalem, which are accessible thanks to the efforts of the Library of Congress (Washington DC).<sup>25</sup> The images may show either one page each (usually in the case of recently produced colour images, see Figures 2–4 above) or two pages side by side; this is typically true of the digitised microfilms (see Fig. 5). The text on the manuscript pages can be arranged in one or two columns so that a maximum of four columns may have to be dealt with in one image if this covers two pages.

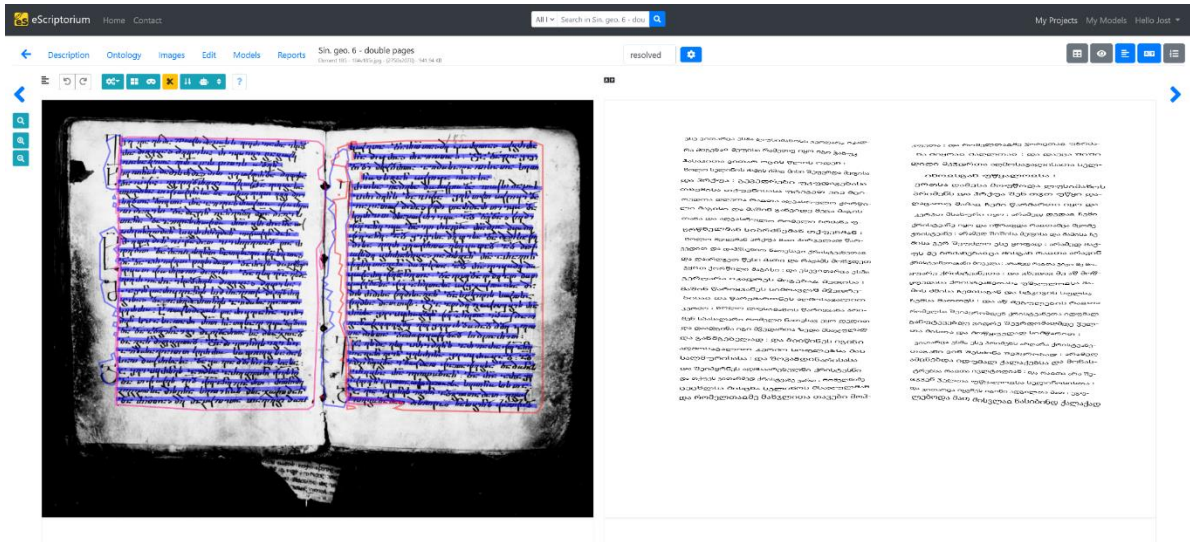


Fig. 5: Double-page microfilm images in *eScriptorium* (fols 184v–185r of Sin. georg. 6)

Old Georgian manuscripts of the first millennium are mostly written in majuscules (*mrglovani*), often with large initial letters (also in majuscules) indicating the beginning of chapters, paragraphs or sections (*asomtavruli*); later manuscripts usually appear in (*nuskha-khutsuri*) minuscules with (*asomtavruli*) majuscules applied to initials, sometimes also titles and other special elements. Whereas the shape of the majuscules is comparatively uniform throughout the given period (roughly the 8th–10th centuries), the minuscules vary to a considerable extent, in both (relative) size and shape and with different degrees of slanting; in addition, the use of abbreviations (usually elisions of one or several characters within a word, indicated by the so-called *karagma*, a tilde-shaped diacritic written above) increases remarkably from century to century. All this means that the development of a unique (one-for-all) solution for a model covering the Old Georgian manuscript production seemed unfeasible right from the beginning; instead, two independent models (one for majuscules only, one for minuscules in combination with majuscules) had to be planned and realised.

### 3.2 A spiral bootstrapping procedure

Since 17 March 2024, four of our project team members<sup>26</sup> have been busy training the two models for Old Georgian.<sup>27</sup> In order to minimise human effort, we have applied a bootstrapping

<sup>25</sup> See <https://www.loc.gov/manuscripts/?q=georgian+jerusalem> and <https://www.loc.gov/manuscripts/?q=georgian+sinai>.

<sup>26</sup> Jost Gippert, Mariam Kamarauli, Eka Kvirkevelia and Sandro Tskhvedadze.

<sup>27</sup> From March to October 2024 we used the msIA server in Paris (<https://msia.escriptorium.fr/>); meanwhile an instance has been set up in Hamburg by Magnus Bender of the Institute for Humanities-Centred Artificial Intelligence (CHAI; see <https://escriptorium.chai.uni-hamburg.de/>).



procedure based on four principles: (1) dividing the manuscripts into two lots according to whether they are written in minuscule or majuscule script, (2) applying “transfer learning” or “fine-tuning” existing segmentation and recognition models for other scripts to Georgian, (3) identifying low-hanging fruit where transcriptions of Georgian manuscripts already exist, and (4) exploiting *eScriptorium*’s automatic alignment feature (text-to-text). Most of this is methodology of Daniel Stökl ben Ezra’s ERC Synergy grant on Hebrew manuscripts, MiDRASH, whose task among others is a transcription of millions of images of manuscripts in Hebrew characters.<sup>28</sup>

We started with three pages of an important minuscule manuscript, the so-called Oshki Bible of 978 CE (manuscript Ivron georg. 1, fols 1r–2r, written in two columns, see Figures 2–4 above), by applying a standard Latin layout segmentation model and correcting it, and then entering the transcription (taken from Akaki Shanidze’s edition of 1947) manually line-by-line. We then fine-tuned a first transcription model optimised for the Oshki Bible on top of the *Biblia* model trained for Medieval Hebrew.<sup>29</sup> On the basis of this tiny dataset, the character accuracy only reached 76.4%, and the word accuracy even only 44.6% after 43 epochs;<sup>30</sup> nevertheless, the model could be used to transcribe subsequent pages automatically and to align existing transcriptions for a first small dataset of five manuscripts (88 pages).<sup>31</sup> For each of them we trained specific models on the same day. On the second day of our “transcribathon” in March, we added text from three further manuscripts<sup>32</sup> and trained a first mixed model for early minuscules, which had a more sizable textual foundation of 116 pages and has since served as a suitable basis for all other minuscule manuscripts we have been dealing with so far.<sup>33</sup> With a similar approach, we also developed the first model for majuscule manuscripts, on the basis of samples from three codices.<sup>34</sup>

We further devised three essential tools for the project to proceed: (1) a tracking sheet (on Google Sheets) where all team members could simultaneously update the status as to which steps had been performed on which part of which manuscript with which model; (2) conventions for the segmentation, and (3) transcription conventions. Right from the beginning we decided not to render the Old Georgian scripts as such in the transcriptions but to transcribe

<sup>28</sup> On an even more automatised methodology applied to Syriac corpora see Bambaci et al. (2024).

<sup>29</sup> Daniel Stökl Ben Ezra (2021): Medieval Hebrew manuscripts version 1.0. (see <https://doi.org/10.5281/zenodo.5468286>).

<sup>30</sup> An “epoch” is a training round in which the computer has seen each element of the training material once.

<sup>31</sup> Single pages, colour, two columns: Oxford, Bodleian Libraries, georg. b1, fols 181r–185v; Tbilisi, Korneli Kekelidze Georgian National Centre of Manuscripts (hereafter: KKNCM), A-95 (the so-called *Parkhali Mravaltavi*), fols 305v–319v; one column: Athos, Ivron Monastery, georg. 45 (autograph by George the Athonite), fols 2v–8v; double pages, black and white, one column: Sinai, St Catherine’s Monastery, georg. 6, fols 184r–200v.

<sup>32</sup> Single pages, colour, two columns: Athos, Ivron Monastery, georg. 10, fols 331v–337v, and georg. 62, fol. 9r; single page, colour, one column: Vienna, Austrian National Library, georg. 2, fol. 1r.

<sup>33</sup> The model is named `georgian_minuscules_02_Ivron_1a_best`.

<sup>34</sup> Single pages, colour, one column: Graz, University Library, ms. 2058/1 (the so-called *Khanmeti Lectionary*), fols 1r–27v; two columns: Athos, Ivron Monastery, georg. 9, fols 1r–3v; Tbilisi, KKNCM, A-1109 (the *Udabno Mravaltavi*), fols 100r–104v. The processing for both models was undertaken by Mariam Kamarauli, Eka Kvirkvelia, and Jost Gippert, with support by Daniel Stökl Ben Ezra.

them into the modern script (*mkhedruli*), in accordance with the practice of scholarly editions of today.

A major challenge consisted in the treatment of the abbreviations (see above). In modern editions of Old Georgian handwritten texts, they are normally expanded. We therefore experimented with three different transcription types: (1) a “diplomatic” rendering of the letters that actually appear in the manuscript, including the *karagma* diacritic; (2) a “resolved” rendering as used in the modern editions; and (3) a “marked” rendering, which is a mixture of both (1) and (2) with the expansion of the abbreviation indicated by parentheses. After some testing, we decided to apply the “resolved” rendering (2) throughout because the system learned those expansions that are not too rare relatively well and because almost all of the existing transcriptions with which we could align contain resolved abbreviations, too.

Another problem we encountered is the division of words between lines. Currently, *passim*, the text-to-text alignment algorithm inside *eScriptorium*, silently splits words between lines if the letters recognised by the rough HTR in each line make this solution more probable. It does, however, not add a hyphen that would indicate the word split at the end of a line as we would find it in modern editions of Old Georgian texts. This makes it more difficult to reconstitute a running text from the automatic transcription. Even if such hyphens, which are practically never graphically represented in any form in the manuscripts, are eventually inserted (manually), the system has only the length and sequence of letters at the end of the given line as data on which to later base its estimation whether to add a hyphen or not. In the current line-based OCR/HTR approach of *kraken*, the contents of a subsequent line are unknown when a given line is processed.

### 3.3 The first results

The two models that resulted from our “transcribathon” of 17 and 20 March 2024 have since then been applied to more than 40 Old Georgian manuscripts, of very different types and sizes, among them the complete Georgian collections of the University Libraries of Graz (manuscripts 2058/1–6c) and Leipzig (manuscripts Vollers 1094–1098), as well as several further manuscripts of the Iviron Monastery on Mount Athos (codices 8, 11, 20, 42 and 89).

The versatility of the models we have developed so far can be illustrated on the example of Leipzig, University Library, Vollers 1095, a bundle comprising, among other fragments, seven pages filled with commemorative notes (ადღაპეობი) from the Georgian community of Jerusalem, written by at least 30 different hands in various inks and styles. After correcting the automatic segmentation, which had yielded an astonishingly correct result in distinguishing lines of extremely unequal shape (see Fig. 6 showing fols 13v–14r), the automatic transcription left no element unread (see Fig. 7), and the alignment with the edition of the notes by Elene Metreveli<sup>35</sup> (see Fig. 8) shows that the reading accuracy was high enough to cover most of the two pages. The few unaligned items could then easily be corrected manually (see Fig. 9).

---

<sup>35</sup> Metreveli (1962: 72–78); cf. the electronic edition in <https://titus.uni-frankfurt.de/texte/etcg/cauc/ageo/liturg/masjer/masje.htm>. A few corrections and additions were provided in Assfalg (1963: 60–72).

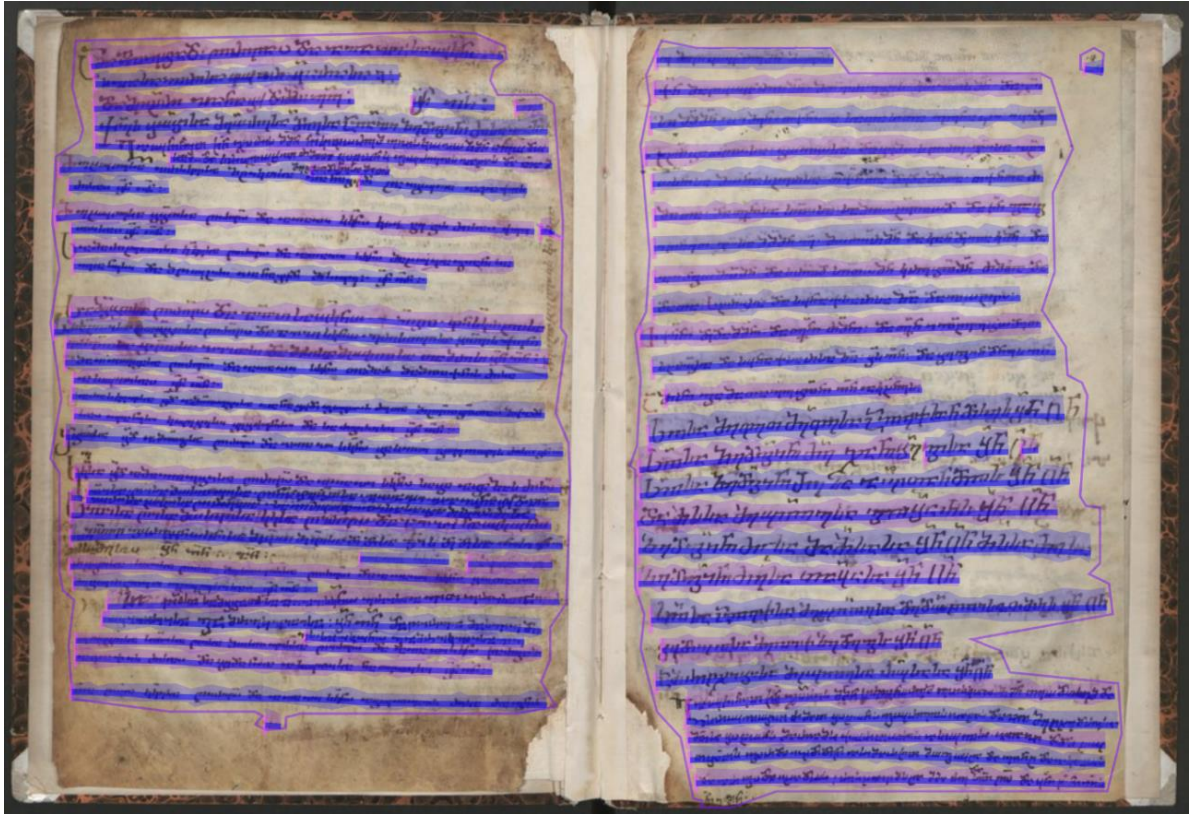


Fig. 6: Leipzig, University Library, Vollers 1095, fols 13v–14r, automatically segmented

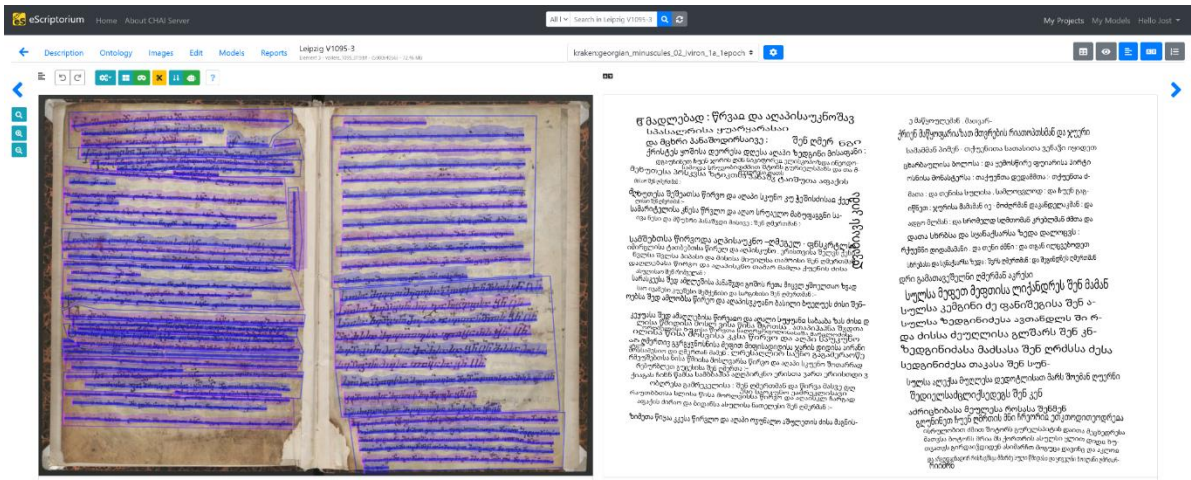


Fig. 7: Same, automatically transcribed after correction of segmentation

As in this case, it has turned out that the automatic segmentation sometimes leads to less satisfying results, mostly due to damages or stains of the writing support, faded inks,<sup>36</sup> erasures, or simply insufficient quality of the images available; this is why we decided to manually correct the segmentation wherever necessary.<sup>37</sup>

<sup>36</sup> Red ink may also pose problems because it appears less dark in the greyscale images into which colour images are converted for the use of the segmentation and transcription algorithms. Another persisting problem is caused by large initials that are outdented and extend over several lines as visible in Figures 11–13 below.

<sup>37</sup> In the manual correction we have been supported by Yorrick Stute and Fahimeh Rahrvan.



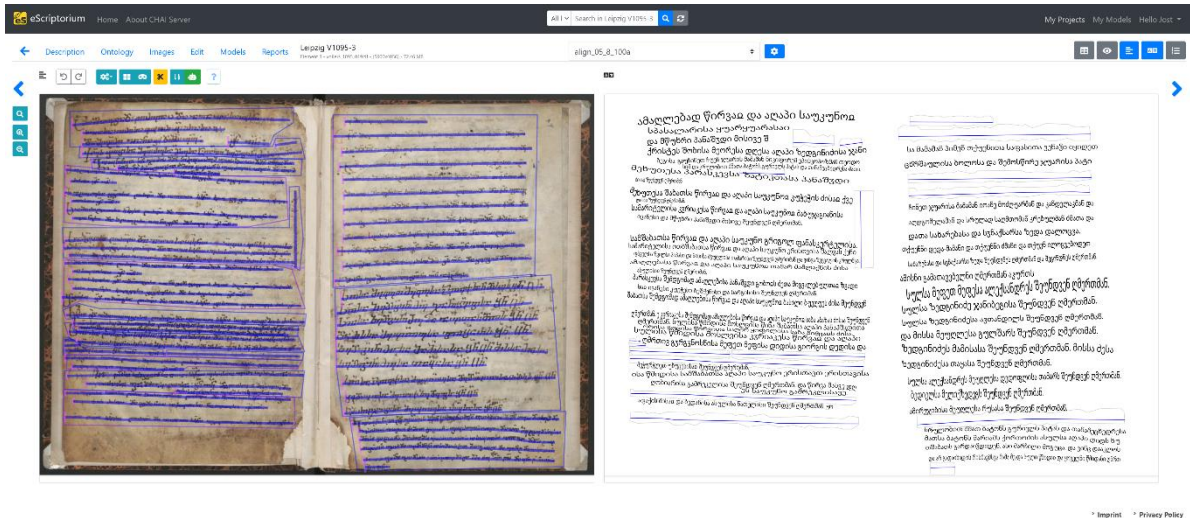


Fig. 8: Same, automatically aligned with existing e-text

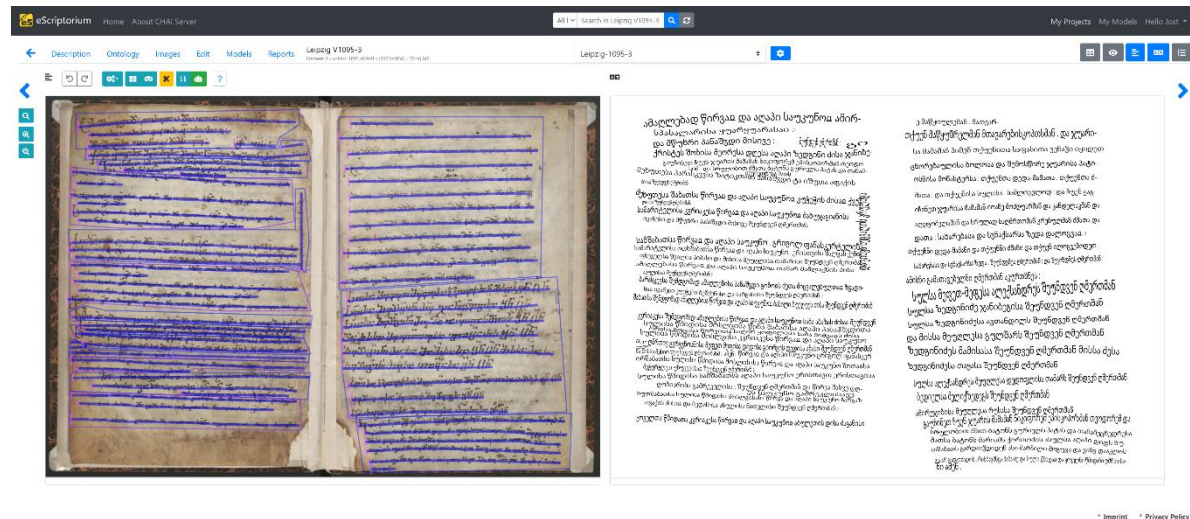


Fig. 9: Same, after manual correction

One of the most problematic cases was manuscript 2058/4 of the Graz collection, a codex written all in majuscules (except for the colophons on fols 94v–95r and 110v, which are in minuscules);<sup>38</sup> Fig. 10 shows fol. 51v after the automatic segmentation and its manual correction. In the given case, the reason for the partial failure of the automatic segmentation seems to be that the image size is smaller than the input size of the neural network for segmentation. After the manual correction, the complete codex was transcribed automatically using the model that was trained for majuscules. The resulting transcription was then aligned with the edition of the text (the liturgy by James and the *Missa Praesantificorum*), which had been provided by Vakhtang Imnaishvili in electronic form for the TITUS project in 2004.<sup>39</sup> As a result, the complete manuscript is now available with near-to full alignment, with only a few lines (less than 6%) needing further manual correction, mostly because they are damaged,

<sup>38</sup> According to the colophon on fol. 95r, the first part of the codex was written by Ioane Zosime on Mount Sinai in 985; the scribe of the second part (fols 96r–110r) was another Iovane (according to his colophon on fol. 110r) who wrote in a totally different hand.

<sup>39</sup> See <https://titus.uni-frankfurt.de/texte/etca/cauc/ageo/liturg/litjak/litja.htm>; the electronic text corresponds to that published in print by Imnaishvili (2004: 266–294).



contain corrections, or abound in abbreviations; cf. Figures 11–13 showing fol. 104v after automatic transcription, automatic alignment, and with the modal window usable for manual correction. Once the aligned text has been corrected in the way indicated and the problem of the insertion of hyphens at the line break has been solved, the transcript thus produced can be used for the development of extended models.

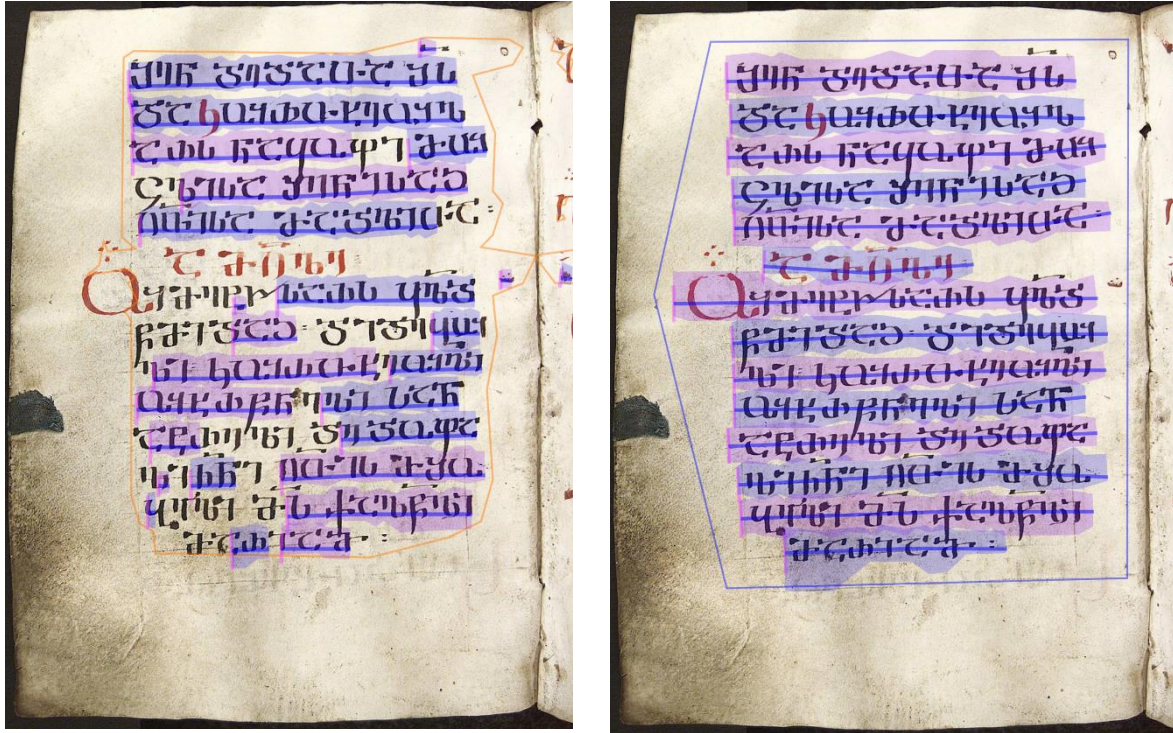


Fig. 10: Segmentation of Graz, University Library, MS 2058/4, fol. 51v (left: automatic; right: manually corrected)

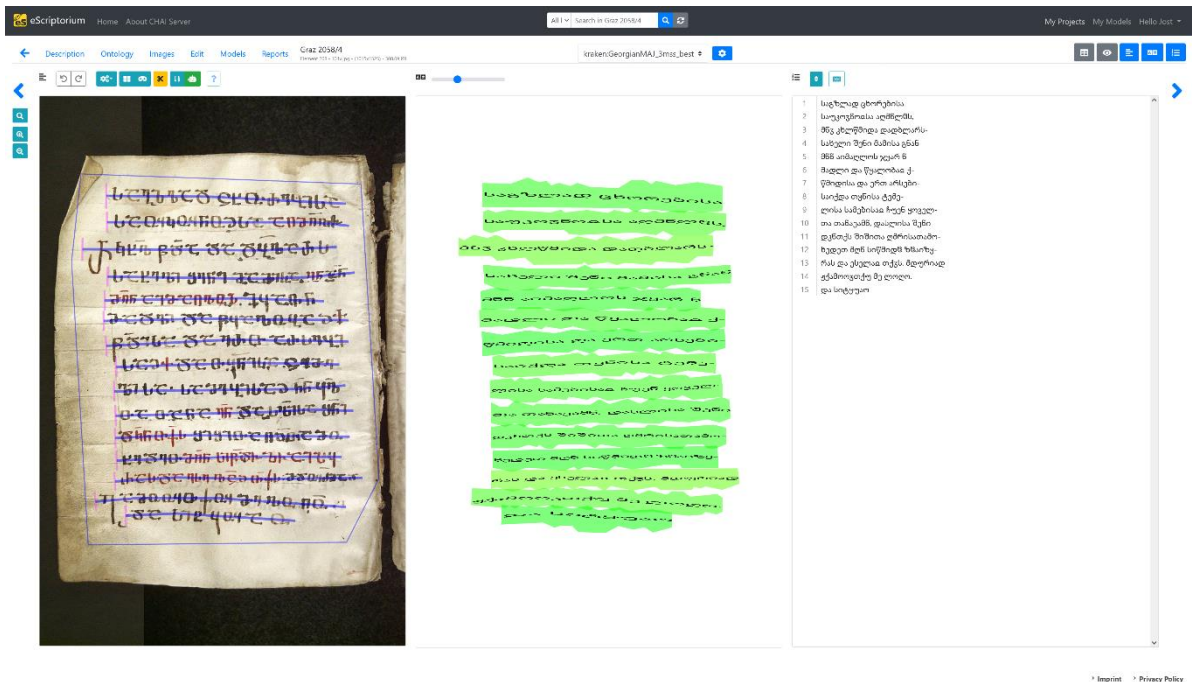


Fig. 11: Graz, University Library, MS 2058/4, fol. 104v, automatically transcribed

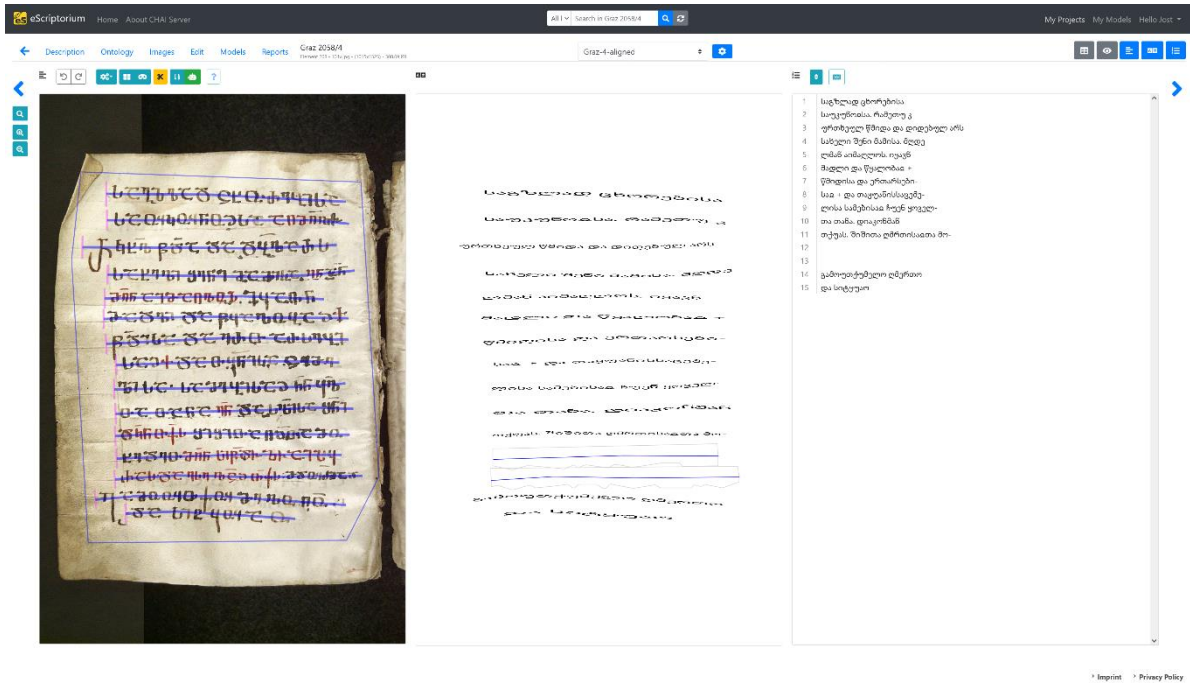


Fig. 12: Same, automatically aligned

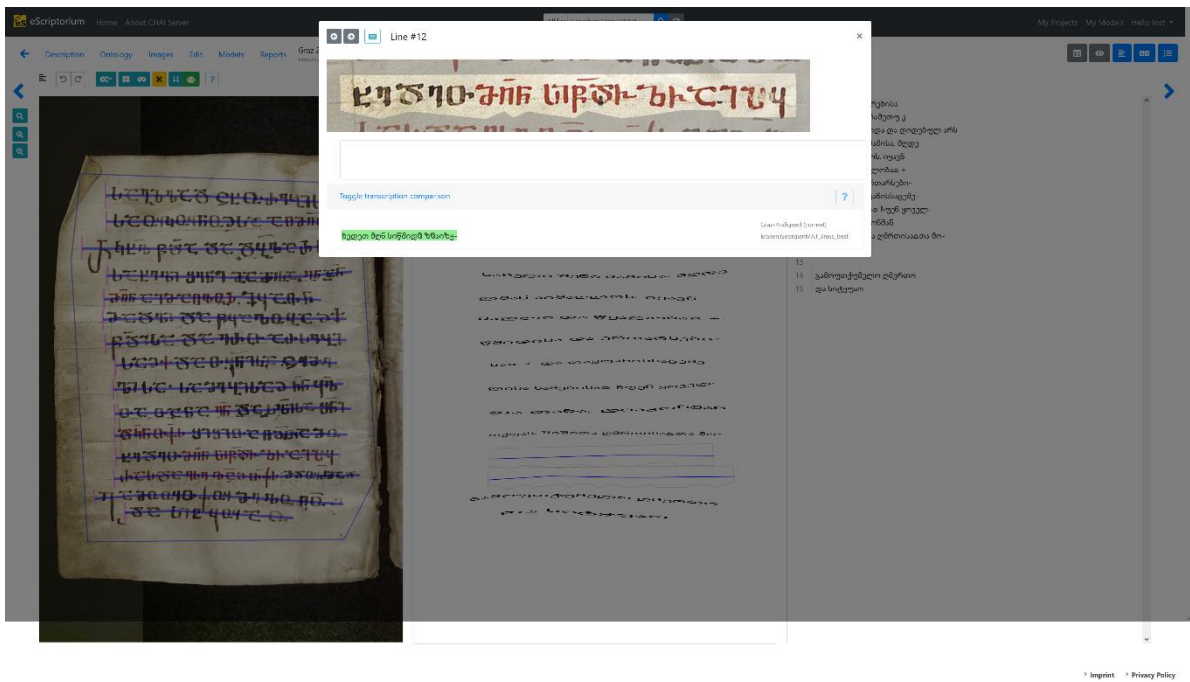


Fig 13: Same, with line 12 shown in modal window

Table II summarises the number of images (= pages) of Georgian manuscripts that have been processed on the *eScriptorium* platform so far (as of 10 November, 2024).

	manually transcribed	auto-aligned	total
majuscules	76	1280	1356
minuscules	216	6208	6424
total	292	7488	7780

### 3.4 Future Tasks

On the basis of the 292 manually transcribed pages and 7488 pages with aligned (but not yet always corrected) text, we are now about to turn towards the next generation of models. We expect these models soon to reach an accuracy reading rate higher than 90%, which is far beyond what we could arrive at with OCR of printed Georgian books 20 years ago. It goes without saying that we intend to make our results openly available; this is not only true of the models but also of the finalised transcripts.<sup>40</sup>

### Acknowledgements

The present research was funded by the European Union’s research and innovation programmes Horizon 2020 (ERC Advanced Grant “DeLiCaTe”, grant agreement no. 101019006) and Horizon Europe (ERC Synergy Grant “MiDRASH”, grant agreement no. 101071829) and by the German Research Foundation (DFG, Cluster of Excellence 2176 ‘Understanding Written Artefacts: Material, Interaction and Transmission in Manuscript Cultures’, project no. 390893796); the work was partly carried out at the Centre for the Study of Manuscript Cultures, University of Hamburg. Views and opinions expressed are those of the authors only and do not necessarily reflect those of the European Union, the European Research Council Executive Agency or the German Research Foundation. Neither the European Union nor the other granting authorities can be held responsible for them.

### References

- Assfalg (1963): Julius A., *Georgische Handschriften* (Verzeichnis der orientalischen Handschriften in Deutschland, 3), Wiesbaden: Steiner. <https://digital.staatsbibliothek-berlin.de/werkansicht/?PPN=PPN1878674927>.
- Bambaci et al. (2024): Luigi B., George Kiraz, Christine Roughan, Matthieu Freyder, Daniel Stökl Ben Ezra, Daniel, “Steps Towards Mining Manuscript Images for Untranscribed Texts: A Case Study From the Syriac Collection at the Vatican Library”, *CHR 2024: Computational Humanities Research Conference*, December 4–6, 2024, Aarhus, Denmark.
- Bonfiglio (2025): Emilio B., “Uncovering Lost Armenian Texts: Schøyen Collection MS 575 and the Armenian Translation of John Chrysostom’s Commentary on the Psalms”, in Gippert, Maksimczuk & Sargsyan (2025), 351–380. <https://doi.org/10.1515/9783111552170-013>.
- Gippert (1990): Jost G., “Perspektiven des Computereinsatzes in der Orientalistik [Perspectives of the application of computers in Near Eastern studies]”, *Forschungsforum* [University of Bamberg] 2, 133–136. <https://titus.fkidg1.uni-frankfurt.de/personal/jg/pdf/jg1990c.pdf>.
- (2023): Jost G., “The Textual heritage of Caucasian Albanian”. In Jost Gippert, Jasmine Dum-Tragut (eds.), *Caucasian Albania. An International Handbook*. Berlin / Boston: de Gruyter, 95–166. <https://doi.org/10.1515/9783110794687-003>.
- (2025a): Jost. G., “Removed and Rewritten: Palimpsests and related phenomena from a cross-cultural perspective”, in Gippert, Maksimczuk & Sargsyan (2025), 1–19. <https://doi.org/10.1515/9783111552170-001>.
- (2025b): Jost G., “Palimpsests from the Caucasus: Two Case Studies”, in Gippert, Maksimczuk & Sargsyan (2025), 253–281. <https://doi.org/10.1515/9783111552170-009>.

---

<sup>40</sup> Whether or not the images can be made available along with the transcripts depends on the institutions owning the manuscripts. We are confident that we will come to the necessary agreements in most cases.

- Gippert, Maksimczuk & Sargsyan (2025): Jost G., José M. and Hasmik S. (eds.), *Palimpsests and Related Phenomena across Languages and Cultures*. Berlin / Boston: de Gruyter (Studies in Manuscript Cultures, 42). <https://doi.org/10.1515/9783111552170>.
- Imnaishvili (2004): ვახტანგ იმნაიშვილი, უძველესი ქართული ხელნაწერები ავსტრიაში. ხანმეტი ლექციონარი, ფსალმუნი, სვიმეონ სალოსის ცხორება, იოვანე ოქროპირისა და იაკობ მოციქულის ჟამისწირვები. თბილისი: ტრიადა.
- Kamarauli (2025): Mariam K., “The Oldest Georgian Witness of the Martyrdom of St Febronia”, in Gippert, Maksimczuk & Sargsyan (2025), 283–308. <https://doi.org/10.1515/9783111552170-010>.
- Kiessling (2022): Benjamin K., *The Kraken OCR system* (Version 4.1.2) [Computer software]. <https://kraken.re>; <https://github.com/mittagessen/kraken>.
- (2024): Benjamin K., *The Kraken OCR system* (Version 5.2.9) [Computer software]. <https://kraken.re>; <https://github.com/mittagessen/kraken>.
- Kiessling et al. (2019): Benjamin K., Robin Tissot, Peter A. Stokes, Daniel Stökl Ben Ezra, “eScriptorium: An Open Source Platform for Historical Document Analysis”, *2019 International Conference on Document Analysis and Recognition Workshops (ICDARW)*, Sydney, NSW, Australia, 2019: 19. <https://doi.org/10.1109/ICDARW.2019.10032>.
- Kvirkvelia (2025): Eka K., “New Witnesses of the Jerusalem-Rite Lectionary: Georgian Palimpsests Ivir. georg. 47 and Ivir. georg. 59”, in Gippert, Maksimczuk & Sargsyan (2025), 309–329. <https://doi.org/10.1515/9783111552170-011>.
- Metreveli (1962): ელენე მეტრეველი, მასალები იერუსალიმის ქართული კოლონიის ისტორიისათვის (XI-XVII სს.), თბილისი: საქართველოს სსრ მეცნიერებათა აკადემია, ხელნაწერთა ინსტიტუტი. <https://dspace.nplg.gov.ge/handle/1234/406705>.
- Mohammed, Jampour & Gippert (2025): Hussein M., Mahdi J. and Jost G., “Inpainting with Generative AI: A Significant Step towards Automatically Deciphering Palimpsests”, in Gippert, Maksimczuk & Sargsyan (2025), 535–545. <https://doi.org/10.1515/9783111552170-019>.
- Sargsyan (2025): Hasmik S., “Linguistic Divergence in Armenian Bible and Lectionary Palimpsests”, in Gippert, Maksimczuk & Sargsyan (2025), 331–350. <https://doi.org/10.1515/9783111552170-012>.
- Shanidze (1947): წიგნნი ძოველისა აღთქოუმისანი 978 წლის ხელნაწერის მიხედვით. ტ. I. ნაკვეთი I. დაბადებისაჲ. გამოსლვათაჲ. გამოსცა აკაკი შანიძემ (ძველი ქართული ენის ძეგლები, 4). თბილისი: საქართველოს სსრ მეცნიერებათა აკადემიის გამომცემლობა.
- Stökl Ben Ezra (2021): Daniel S., Medieval Hebrew manuscripts version 1.0. *Zenodo*. <https://doi.org/10.5281/zenodo.5468286>.
- (forthcoming): Daniel S., “Text Digitization”. To appear in Christopher A. Nunn and Frederike van Oorschot (eds.), *Compendium of Computational Theology 1. Introducing Digital Humanities to Theology*, Heidelberg: heiBOOKS (probably 2024), 99–115.
- Stökl Ben Ezra et al. (2021): Daniel S., Bronson Brown-DeVost, Pawel Jablonski, Hayim Lapin, Benjamin Kiessling, Elena Lolli, “BibLIA – a General Model for Medieval Hebrew Manuscripts and an Open Annotated Dataset”. *HIP@ICDAR 2021*: 61–66. <https://doi.org/10.1145/3476887.3476896>.
- Stokes et al. (2021): Peter Anthony S., Benjamin Kiessling, Daniel Stökl Ben Ezra, Robin Tissot, El Hassane Gargem, “The eScriptorium VRE for Manuscript Cultures”, *Classics@*, 2021, 18 (1). <https://classics-at.chs.harvard.edu/classics18-stokes-kiessling-stokl-ben-ezra-tissot-gargem/>.



# ქართული ხელნაწერების ავტომატური ამოკითხვა *eScriptorium*-ის პლატფორმაზე

იოსტ გიპერტი (ჰამბურგი), დანიელ შტეკელ ბენ ეზრა (პარიზი)

DOI: <https://doi.org/10.62235/dk.3.2024.8508>

[jost.gippert@uni-hamburg.de](mailto:jost.gippert@uni-hamburg.de) || ORCID: [0000-0002-2954-340X](https://orcid.org/0000-0002-2954-340X)

[Daniel.Stoekl@ephe.psl.eu](mailto:Daniel.Stoekl@ephe.psl.eu) || ORCID: [0000-0001-5668-493X](https://orcid.org/0000-0001-5668-493X)

ჰუმანიტარიის არსებობის ისტორიის განმავლობაში არაფერს არ შეუცვლია კვლევის პროცესი ენათმეცნიერებაში, ფილოლოგიასა და მომიჯნავე დარგებში ისე მკვეთრად, როგორც ეს ციფრულ ეპოქაში მოხდა: დიდ მონაცემთა ტექსტური ბაზების შექმნამ, კორპუსული კვლევების განვითარებამ და ავტომატური ძიების მრავალფეროვანმა მეთოდებმა სწრაფი და ეფექტიანი კვლევის საშუალება მისცა მკვლევრებს. ტექსტური კორპუსების განხილვისას უნდა განვასხვაოთ ორი ტიპის ტექსტური რესურსები: „ციფრულად დაბადებული“ (ანუ ის რესურსები, რომლებიც თავიდანვე ელექტრონული ფორმით შეიქმნა) და რეტროდიგიტალური რესურსები (ხელნაწერი ან ბეჭდური სახით არსებული რესურსები, რომლებიც გაციფრებას საჭიროებს). ქართული ენის ეროვნული კორპუსი, რომელიც ბოლო 13 წლის განმავლობაში შეიქმნა (GNC, <http://gnc.gov.ge>), ორივე ტიპის რესურსებს მოიცავს: მისი ყველაზე დიდი ქვეკორპუსი (ქართული ენის რეფერენციული კორპუსი [GRC], რომელიც 202 728 329 ტოკენის მოცულობისაა) ძირითადად შედგება იმ მასალებისგან, რომლებიც შეგროვდა WWW (World Wide Web) ქსელში; ხოლო ქვეკორპუსები, განსაკუთრებით ისინი, რომლებიც ძველი (GNC-ის ძველი ქართულის ქვეკორპუსი, 7 101 021 ტოკენი) და საშუალო ქართული ენის (GNC-ის საშუალო ქართულის ქვეკორპუსი, 1 432 262 ტოკენი ) რესურსებს მოიცავს, ასევე თემატური კორპუსები, როგორებიცაა ქართული სამართლის ტექსტები (GNC-ის იურიდიული ტექსტების ქვეკორპუსი, 1 495 985 ტოკენი) ან პოლიტიკური ტექსტები (GNC-ის პოლიტიკური ტექსტების ქვეკორპუსი, 1 436 075 ტოკენი) მეტ-ნაკლებად ეფუძნება ბეჭდური მასალის გაციფრებულ რესურსებს. ისინი TITUS-ისა და ARMAZI-ის პროექტების ფარგლებში შეიქმნა. რესურსების მხოლოდ მცირე ნაწილი გაციფრულდა ხელით, ანუ მოცემული ტექსტის კლავიატურაზე აკრეფით; უმეტეს შემთხვევაში დაბეჭდილი ტექსტები მუშავდებოდა ოპტიკური სკანერის მეშვეობით და შემდეგ „იკითხებოდა“ ციფრული ტექსტის ფორმატში სპეციალური პროგრამული უზრუნველყოფის საშუალებით, რომელსაც შეეძლო გრაფემების განცალკევება, მათი ამოცნობა ანბანის შესაბამის სიმბოლოებთან შედარების გზით და ამოცნობილი ასოების თანმიმდევრულად შენახვა ელექტრონული ტექსტის ფორმატში.

1980-იანი წლების ბოლოს, როდესაც საფუძველი ეყრებოდა ქართული ენის დიაქრონიული კორპუსის შექმნას, სიმბოლოების ოპტიკური ამოცნობის (OCR) პროგრამები ჯერ კიდევ არ იყო ისე კარგად განვითარებული, როგორც დღეს არის, და სკანერებიც ტექნიკურად ბევრად ჩამორჩებოდნენ თანამედროვე სკანერებს. თუმცა ყველაზე დიდი პრობლემა მაინც იმაში მდგომარეობდა, რომ იმდროინდელი პერსონალური კომპიუტერები არ იყო გათვლილი იმ ენებზე სამუშაოდ, რომელთაც ლათინურისაგან

განსხვავებული ანბანები აქვთ. ისინი 7-ბიტის კოდირების სისტემაზე იყო დაფუძნებული (ASCII) და მხოლოდ ინგლისურ ენაში გამოყენებულ სიმბოლოებს მოიცავდა. თანდათანობით 7-ბიტის სისტემა შეიცვალა 8-ბიტის სისტემით (ANSI), რომელიც უკვე ფარავდა „დამატებით სიმბოლოებს“, რომლებიც აუცილებელი იყო „დასავლური“ ენებისთვის, მაგალითად, გერმანულისა („უმლაუტის სიმბოლოები“) ან ფრანგულისთვის („მახვილებიანი სიმბოლოები“). საერთაშორისო სტანდარტის ISO/IEC-8859 დანერგვით ლათინურზე დაფუძნებული „ჩრდილო“, „ცენტრალური“ და „სამხრეთ ევროპული“ ანბანების გვერდით შესაძლებელი გახდა კირილიცასა და ბერძნული ანბანის გამოყენება. მათგან განსხვავებით ქართულ ანბანს იმ პერიოდში საერთოდ არ ჰქონდა კომპიუტერული მხარდაჭერა, რაც საგრძნობლად აფერხებდა ქართული ენის გაციფრულებას. მიუხედავად მეცნიერთა და ინფორმატიკის დარგის სპეციალისტთა ძალისხმევისა, რომლებმაც რამდენიმე შუალედური ოპტიკური ხელსაწყო შექმნეს, არსებითად არ შეცვლილა ქართული ენის გაციფრულებასთან დაკავშირებული ტექნიკური სირთულეები.

მოგვიანებით, 1990-იან წლებში, ბაზარზე გამოჩნდა ახალი ოპტიკური პროგრამა „FineReader“, რომელმაც გადამწვევტი როლი შეასრულა ქართული ტექსტების მასშტაბურ გაციფრულებაში. მიუხედავად იმისა, რომ ქართული არ იყო იმ მრავალ ენათა რიცხვში, რომელთაც „FineReader“ უჭერდა მხარს, მაინც მოხერხდა მისი „გავარჯიშება“ ქართული ენის ტექსტების წასაკითხად. TITUS-ისა და ARMAZI-ის პროექტების ფარგლებში გაციფრულებული ძველი და საშუალო ქართულის რესურსების უმეტესი ნაწილი სწორედ მისი გამოყენებით შეიქმნა. 1990-იანი წლების ბოლოს კი დადგა გადამწვევტი ეტაპი – შემუშავდა უნიკოდის სტანდარტი (ISO/IEC-10646), რომელიც მოიცავდა უნიკალური კოდირების სისტემას ქართული ანბანის სამივე სახეობისათვის (მხედრული, ნუსხა-ხუცური და ასომთავრული/მრგლოვანი).

თანამედროვე, ციფრული საუკუნის პირობებში მიმდინარე ტექნიკურმა რევოლუციამ ოპტიკური ამოცნობის სრულიად ახალი პერსპექტივები შექმნა, რამაც შესაძლებელი გახადა აქამდე გადაუჭრელი არაერთი პრობლემის დაძლევა, განსაკუთრებით ხელნაწერების გაციფრულების თვალსაზრისით. დღეისათვის შექმნილია ხელოვნურ ინტელექტზე (AI) დაფუძნებული ხელნაწერი ტექსტის ამოცნობის (HTR) წვრთნადი პროგრამული უზრუნველყოფა. პროექტის „წიგნობრობის განვითარება კავკასიის ტერიტორიებზე (DeLiCaTe)“ ფარგლებში ამ სისტემის ქართულ ხელნაწერებზე გამოყენების პირველმა ცდამ ბოლო ექვსი თვის განმავლობაში შესანიშნავი შედეგი გამოიღო. ხელნაწერი ტექსტების ამოცნობის პლატფორმა *eScriptorium* შემუშავებულია პარიზის უნივერსიტეტში (Paris Sciences et Lettres University) სამეცნიერო პროექტების *Scripta* და *RESILIENCE* ფარგლებში, რომელთა განხორციელებაში ჩართული იყო სხვა ინსტიტუტებიც. პროექტები ნაწილობრივ დაფინანსდა ევროკავშირის პროგრამის „ჰორიზონტი 2020“-ისა და ენდრიუ ვ. მელონის ფონდის (Andrew W. Mellon Foundation) მიერ.

1980-იანი წლების ბოლოდან ქართული ბეჭდური წიგნების გაციფრულებისთვის სიმბოლოების ოპტიკური ამოცნობის (OCR) ხელსაწყოების დასახეწად გაწეული ძალისხმევის მიმოხილვის, აგრეთვე, ხელნაწერი ტექსტის ამოცნობისა (HTR) და *eScriptorium*-ის შესახებ მოკლე შესავლის შემდეგ სტატიაში მოცემულია ამ უკანასკნელის ფუნქციონირების ძირითადი ასპექტები. ქართული ხელნაწერების ავტომატურად გადმოწერის ხარისხის გაუმჯობესების მიზნით *eScriptorium*-ის წვრთნა მოიცავს რამდენიმე

ძირითად ეტაპს: პლატფორმაზე ხელნაწერის ციფრული ფოტოების ატვირთვას, თითოეულ გვერდზე ტექსტით დაფარული სვეტ(ებ)ისა და სტრიქონების ავტომატურ სეგმენტაციას (სეგმენტაციის კორექტირება ზოგჯერ ხელით ხდება), მონიშნული გვერდების ავტომატურ ტრანსკრიბირებასა და ტრანსკრიბირებული ტექსტის ავტომატურ გასწორებას („alignment“) ხელნაწერის უკვე არსებული გამოცემის ტექსტთან. ზოგიერთი ხელნაწერის შემთხვევაში ტრანსკრიბირება მოხდა ხელით (ასეთი ამ ეტაპისთვის 292 გვერდია). რაც შეეხება ავტომატურად ტრანსკრიბირებულსა და გასწორებულს არსებული გამოცემის ტექსტთან, მათი ოდენობა 7488 გვერდს შეადგენს. დაგროვილი მასალა საშუალებას გვაძლევს, გავაუმჯობესოთ ტრანსკრიბირების აქამდე არსებული მოდელი. ველით, რომ ხელნაწერის ამოკითხვის სიზუსტის მაჩვენებელი 90%-ს გადააჭარბებს. აღსანიშნავია ისიც, რომ დაინტერესებული საზოგადოებისათვის ჩვენი შედეგები სრულად ხელმისაწვდომი იქნება. ეს ეხება როგორც უშუალოდ ტრანსკრიბირების მოდელს, ასევე, გამოუცემელ ხელნაწერთაგან გადმოწერილ და გასწორებულ ტექსტებსაც.

# Digital Humanities and the Korneli Kekelidze Georgian National Center of Manuscripts

Maia Matchavariani (Tbilisi)

DOI: <https://doi.org/10.62235/dk.3.2024.8509>

[matchavarianimn@yahoo.com](mailto:matchavarianimn@yahoo.com) || ORCID: [0009-0008-7155-1068](https://orcid.org/0009-0008-7155-1068)

**Abstract:** The recognition of Digital Humanities as a scientific discipline emerged at the K. Kekelidze Institute of Manuscripts by the late 20th century. In the early 1990s, a first project in the field of Digital Humanities was initiated – the development of an automated database of Georgian historical documents. The database was designed to facilitate access to these invaluable resources for both Georgian and international scholars while contributing to the broader mission of promoting Georgia’s cultural heritage. In addition to cataloguing and preserving historical documents, the project had a broader, strategic goal, namely, to illustrate the transformative impact of emerging technologies on Humanities research. It aimed to foster a paradigm shift by encouraging scholars to adopt digital tools as integral part of modern research methodologies. Since then, the Institute – later renamed the Korneli Kekelidze Georgian National Center of Manuscripts – has continued to lead numerous digitally-based projects supported by the Rustaveli Scientific Foundation and in collaboration with various academic institutions. Some of these initiatives are outlined in the present paper.

**Keywords:** Digital Humanities, Manuscripts, Korneli Kekelidze Georgian National Center of Manuscripts

## Introduction

The significance of Digital Humanities as a scientific discipline was recognized at the K. Kekelidze Institute of Manuscripts by the late 20<sup>th</sup> century. In the early 1990s, the introduction of the first computers at the Institute, under the guidance of Elene Metreveli, led to the creation of the Department of Scientific Information and Informatics, which embarked on pioneering Digital Humanities projects. The initial project aimed to develop an automated database of Georgian historical documents, with a preliminary description published in 1991 in the Institute’s periodical *Mravaltavi*.<sup>1</sup> This database was designed not only to provide easier access to these invaluable materials for both Georgian and international scholars but also to play a crucial role in the broader effort to popularise Georgia’s cultural heritage.

Beyond the immediate objective of cataloguing and preserving historical documents, the project had another, more strategic aim: to demonstrate the transformative potential of new technologies in the Humanities. It sought to advocate for a paradigm shift in research methodologies, encouraging scholars to embrace digital tools as essential components of modern Humanities research. Since then, the Institute, later renamed as the Korneli Kekelidze Georgian National Center of Manuscripts, has undertaken numerous projects with the support of the Shota Rustaveli National Science Foundation of Georgia and in collaboration with various academic institutions. Some of these projects are presented below.

---

<sup>1</sup> Abashidze & Surguladze 1991.



## **1. Georgian Script and Paleographic Album (2012–2014)**

Project leaders: Zaza Aleksidze, Buba Kudava

Project group: Zaza Aleksidze, Ketevan Asatiani, Buba Kudava, Elene Machavariani, Tamar Otkhmezuri, Teimuraz Jojua, Dali Chitunashvili

The *Georgian Script and Paleographic Album* project (2012–2014) aimed to address the challenge of dating undated Georgian manuscripts by creating a paleographic album and providing scholars with a systematic tool for manuscript dating. Key features and achievements include:

- Focus on dating Georgian manuscripts: Since only a small portion of Georgian manuscripts contain explicit dates, the project aimed to develop methodologies for dating undated manuscripts through paleographic analysis, which involves studying the style and evolution of handwritten scripts.
- Global search for dated manuscripts: The research team conducted a comprehensive search of repositories worldwide to identify and analyse Georgian manuscripts with known dates. These manuscripts provided the necessary data to create a robust, reliable paleographic database.
- Development of a paleographic database: The project resulted in the creation of a paleographic database which systematically catalogues the scripts of dated Georgian manuscripts. This database is an invaluable resource for paleographers, historians, and manuscript scholars, providing a structured way to analyse the evolution of Georgian script and draw comparisons between dated and undated texts.
- Publication of results: The findings of the project were published in a paleographic album based on Georgian manuscripts with known dates.<sup>2</sup> The album provides detailed insights into the methodology and results of the paleographic study. This book, with the paleographic album contained in it, forms a foundational resource for further research in Georgian manuscript studies.
- Advancing Georgian paleography: The project made a significant contribution to the field of Georgian paleography by offering a scientific tool that enhances the ability to date undated manuscripts, which is critical for establishing their historical context and provenance.

Through these efforts, the project has enriched the study of Georgian manuscripts, providing essential resources for future scholarship.

## **2. Georgian, Persian and Ottoman Illuminated Documents from the Depositories of Georgia (Database, Website, Album) (2007–2011)**

Project leader: Darejan Kldiashvili

Project Group: Darejan Kldiashvili, Tamar Abuladze, Tsisana Abuladze, Irina Koshoridze (Georgian National Museum), Zaza Skhirtladze (Tbilisi State University)

---

<sup>2</sup> Aleksidze et al. 2015.

The project *Georgian, Persian, and Ottoman Illuminated Documents from the Depositories of Georgia* (2007–2011) aimed to digitise and study illuminated documents preserved in Georgia, dating from the 11<sup>th</sup> to 19<sup>th</sup> centuries. Conducted by the National Center of Manuscripts, the project focused on examining the artistic, historical, and cultural significance of these documents, written in Georgian and other languages. Below are the key features and achievements of the project:

- Scope of research: The project studied and digitised over 200 illuminated documents, including royal court documents, firmans, dowry books, and legal records from both Georgian and Persian sources. These documents span from the 11<sup>th</sup> to 19<sup>th</sup> century, providing a comprehensive look at the artistic and cultural heritage of Georgia and its interaction with neighbouring regions.
- Chronological mapping of illuminations: The project traced the artistic development of Georgian illuminated documents. Early Georgian documents (11<sup>th</sup>–13<sup>th</sup> centuries) focused on material, form, and calligraphy, and it was only in the 14<sup>th</sup>–15<sup>th</sup> centuries that miniatures began to appear, often featuring portraits of historical and royal figures. This evolution highlighted the growing artistic sophistication over time.
- Artistic influence and independence: Georgian illuminated documents retained a unique artistic identity, even as the country became more integrated into the Islamic world after the fall of Byzantium. Despite political and cultural changes, these documents reflected an independent tradition, combining Georgian, Byzantine, and Islamic decorative influences.
- Multicultural focus: In addition to Georgian documents, the project extensively examined Persian and Ottoman documents. These materials, such as firmans and dowry books, were issued by Christian and Muslim rulers of Georgia, as well as the semi-independent khanates of Yerevan and Dagestan. This provided insight into the multicultural political landscape and the artistic blend of Christian and Islamic traditions.
- Digital archiving: The project successfully digitised the illuminated documents, making them accessible through a publicly available online database ([illuminateddocument.ge](http://illuminateddocument.ge)). This digital resource provides high-quality images and detailed descriptions of the manuscripts, allowing for wider access and further research.
- Publication of findings: The results of the project were published in a book which includes 250 black-and-white and 27 colour photographs of the documents, along with scholarly commentaries contributing to the academic study of illuminated manuscripts.<sup>3</sup>
- Interdisciplinary analysis: The documents were studied from multiple perspectives – codicological, historical, and artistic –, thus offering a well-rounded understanding of their significance. The project shed light on the region's illuminated manuscript tradition and its connections with the neighbouring cultural and artistic traditions from Byzantium to Persia.

---

<sup>3</sup> Kldiashvili 2011.

- Preservation of heritage: By documenting and analysing these manuscripts, the project helped preserve Georgia's illuminated document tradition, highlighting the region's contributions to the wider field of art history and manuscript studies.

This project has deepened our understanding of the cultural and artistic exchanges between Georgia, Persia, Byzantium, and the Ottoman Empire, preserving valuable heritage for future generations.

### **3. Astronomical Manuscripts in Georgia (2012–2015)**

Project leader: Irakli Samsonia (Ilia State University)

Project group: Irakli Samsonia (Ilia State University), Tamar Abuladze, Liana Samkurashvili, Nestan Chkhikvadze

The *Astronomical Manuscripts in Georgia* project (2012–2015), led by the National Center of Manuscripts in collaboration with Ilia State University (Tbilisi, Georgia), aimed to research, catalogue, and systematise astronomical manuscripts preserved in Georgia. This scholarly initiative focused on manuscripts written in Georgian, Persian, and Arabic, dating from the 10<sup>th</sup> to the 19<sup>th</sup> centuries. Here are the key features and achievements:

- Comprehensive study of 360 manuscripts: The project involved the examination of 360 manuscripts, shedding light on Georgia's rich astronomical tradition, as well as its connections to Persian and Arabic scientific cultures.
- Creation of an electronic database: A key outcome was the development of a comprehensive online database and webpage, providing detailed information about each manuscript. The database includes annotations and visual materials sourced from Georgian and international collections (Mount Athos, Mount Sinai, St. Petersburg). This resource also features a dictionary of astronomical terms, significantly aiding researchers in the history of science.
- Manuscript classification: The project categorised Georgian manuscripts into four main groups:
  - a) Purely astronomical research and descriptions.
  - b) Notes on the use of specific methods, instruments, and results.
  - c) Textbooks and instructional materials.
  - d) General educational texts. Many of these manuscripts were translations from Greek, Persian, Arabic and other sources, reflecting Georgia's engagement with neighbouring scientific knowledge.
- Unstudied astronomical content: The scientific content of the Georgian manuscripts was closely linked to ancient cosmological ideas, calendar-chronological systems, and spherical astronomy. The study revealed previously unexamined material, offering fresh insights into the cultural and scientific exchanges between Georgia, Byzantium, and Persia. These findings contribute to the understanding of Georgia's role in the broader narrative of intellectual history in the region.

- In-depth analysis of Persian and Arabic manuscripts: In addition to Georgian texts, the project included a detailed analysis of Persian and Arabic manuscripts housed at the National Center of Manuscripts. This collection included copies of classical works by oriental astronomers and mathematicians, along with lesser-known manuscripts covering cosmology, spherical astronomy, planetary and stellar descriptions, calendar systems, and star catalogues.
- Fostering cross-cultural understanding: The project emphasised the intricate connections between Georgian, Byzantine, Persian, and Arabic scientific traditions. By cataloguing and digitising these manuscripts, the project enhanced the accessibility of these documents for future research in the fields of the history of science, astronomy, and manuscript studies.

Overall, the *Astronomical Manuscripts in Georgia* project made important contributions by preserving and opening access to a rich body of astronomical knowledge. It highlighted Georgia's active role in the scientific exchanges between different cultures and provided invaluable resources for scholars exploring the history of astronomy.

#### **4. Digitalization and Database of Historical Documents (2012–2014)**

Project leader: Tea Kartvelishvili

Project group: Tea Kartvelishvili, Nikoloz Zhghenti, Tamaz Gogoladze, Irakli Gelashvili, Maia Shaorshadze, Nestan Bagauri, Otar Zhizhiashvili

The *Digitalization and Database of Historical Documents* project, carried out between 2012 and 2014 by the National Center of Manuscripts, was an important initiative aimed at preserving and providing access to Georgia's vast collection of historical documents. The Center safeguards over 40,000 historical documents, which span nearly every aspect of Georgian life during the Middle Ages, from royal decrees to personal letters and records of household relations. Given the originality and reliability of these documents, their preservation and accessibility are of paramount importance to scholars and researchers. The project addressed these needs by undertaking the following key tasks:

- Selection and digitisation: A total of 5,000 historical documents were selected for digitisation. The selection criteria included the age of the documents, the extent of physical damage, and the significance of their content. These documents, representing a broad cross-section of Georgian history, were digitised to ensure their preservation and to make them more easily accessible to researchers.
- Creation of a database: The digitised documents were uploaded to a specially designed database. This database is more than just a digital repository; it also includes detailed archeographic descriptions of each document. These descriptions provide essential information, such as:
  - the type of the document (e.g., decree, letter, legal record)
  - a brief summary of the content
  - the exact or approximate date of the document
  - the place of creation



- historical figures mentioned in the document
- geographical locations referenced
- details about seals and signatures present.

This comprehensive metadata allows researchers to search for information across various criteria, such as by date, historical figures, or geographical area. Even within the scope of 5,000 selected documents, this search functionality provides a rich and detailed picture of specific historical events, relationships, and contexts. By digitising and systematically cataloguing these documents, the project has ensured that they can be studied in greater depth and with greater ease, helping to foster a deeper understanding of Georgia's historical and cultural legacy.

## **5. Georgian Manuscripts Preserved in the Regions of Georgia: Unified Electronic Database (in Georgian and English Languages) (2013–2015)**

Project leader: Shalva Gloveli

Project group: Nestan Chkhikvadze, Grigol Gagnidze (Georgian National Museum), Shalva Gloveli, Darejan Gogashvili, Maia Karanadze, Tamar Otkhmezuri

The project *Georgian Manuscripts Preserved in the Regions of Georgia: Unified Electronic Database (in Georgian and English Languages)*, carried out between 2013 and 2015 in collaboration with the Georgian National Museum, was an important initiative aimed at cataloguing and making accessible the Georgian manuscripts dispersed throughout the country. The primary goal of the project was to create a unified, bilingual (Georgian-English) electronic catalogue and database of these manuscripts, while also revising and updating unpublished catalogues of regional manuscripts that were originally compiled in the 1960s and 1970s.

The project involved an extensive survey and documentation process, covering a wide range of institutions and private collections across Georgia, including regional museums, house museums, libraries, archives, churches, monasteries, mosques, synagogues, and family collections. The database created through this project includes detailed information on all types of manuscripts, from complete works to those with missing parts and fragments. The manuscripts documented in the regions are diverse in content, encompassing religious texts, literary works, historical chronicles, medical treatises, astrological writings, philosophical texts, legal documents, and more. The key features and achievements of the project are as follows:

- Unified catalogue and bilingual database: The project compiled a comprehensive, bilingual (Georgian-English) electronic catalogue and database that provides essential information about Georgian manuscripts preserved in various regions. This resource is invaluable for scholars and researchers, facilitating access to previously inaccessible or little-known materials.
- Publication of findings: The results of the project were published in a bilingual catalogue which includes information on 1,305 Georgian manuscripts kept in the regions of the country.<sup>4</sup>

---

<sup>4</sup> Gagnidze et al. 2015.

- Consultations and corrections: During the project, significant consultations were conducted with regional museums and other institutions. These consultations led to the correction of catalogue entries, manuscript titles, and the dating of manuscript covers, ensuring the accuracy and reliability of the database.
- Discovery and documentation: One of the major outcomes of the project was the identification and documentation of a large number of manuscripts that had not been previously known to the scholarly community. By bringing these manuscripts to light, the project has greatly expanded the corpus of Georgian manuscript heritage available for study.
- Inclusion of archival materials and early printed books: In addition to manuscripts, the project also gathered information on related archival materials, historical documents, and early printed books found in the regions. While these materials are not yet fully integrated into the database, they are currently being prepared for inclusion, which will further enhance the resource.

By creating a centralised bilingual database, the project has provided scholars both within Georgia and internationally with a powerful tool for research, helping to ensure that these important cultural artifacts are recognised and studied in greater detail. Moreover, the project has strengthened the connection between regional institutions and the broader academic community, fostering a more comprehensive understanding of Georgia's rich manuscript tradition.

## **6. Database of Miniatures Involved in the Decoration of Georgian Illustrated Manuscripts (9<sup>th</sup>–19<sup>th</sup> centuries) (2014–2016)**

Project leader: Ketevan Tatishvili

Project group: Ketevan Tatishvili, Elene Machavariani, Nino Tsitsishvili

The project *Database of Miniatures Involved in the Decoration of Georgian Illustrated Manuscripts (9<sup>th</sup>–19<sup>th</sup> cc.)*, conducted between 2014 and 2016, was an extensive initiative aimed at cataloguing and providing comprehensive access to the rich tradition of Georgian manuscript illumination. The primary objective was to create a bilingual database that would serve as a detailed iconographic index of Georgian illustrated manuscripts from the 9<sup>th</sup> to the 19<sup>th</sup> centuries, offering scholars and researchers a valuable resource for studying the artistic and contextual aspects of these manuscripts. The key features and achievements of the project are as follows:

- Creation of a bilingual database: A central outcome of the project was the development of a bilingual (Georgian-English) database that allows users to search for specific material and illustrations using defined classifiers. This database is a significant scholarly tool, offering detailed information about both the manuscripts and the individual miniatures they contain.
- Comprehensive manuscript information: The database includes extensive details about each manuscript, such as the content, date, and place of copying, as well as the dimensions, number of pages, ink, writing material, text distribution, and information

about the artist, copyist, engraver, and cover maker. Additionally, the database provides descriptions of the artistic decoration, colophons (notes), and relevant bibliographic references, making it a thorough resource for manuscript studies.

- Detailed miniature descriptions: Each miniature within the database is meticulously described with information including its name, page location, size, placement within the manuscript, and the characters, objects, flora, and fauna depicted. The database also notes any other characteristic details of the decoration. This level of detail allows for a deep understanding of the iconography and artistic conventions of Georgian manuscript illumination.
- Focus on illumination principles: One of the unique merits of the database is its focus on the illumination principles of Georgian manuscripts as indivisible codices. The project emphasises the unity of contextual and artistic-decorative aspects by linking miniatures to relevant textual captions from the codices in which they are found. This approach ensures that the miniatures are understood not just as isolated artworks but as integral parts of the manuscripts' overall design and meaning.
- Inclusion of manuscripts from various depositories: While the project primarily focuses on illustrated manuscripts preserved at the National Center of Manuscripts, it also includes artistically decorated manuscripts from other depositories across Georgia. This inclusion broadens the scope of the database, making it a more comprehensive resource for the study of Georgian manuscript illumination.

By providing a unified platform for the study of these miniatures, the project has created a critical resource for understanding the interplay between text and image in Georgian manuscripts, ensuring that this important aspect of Georgian heritage is preserved and studied in depth.

## **7. Natural Disasters in the South Caucasus: Historical Earthquakes and Associated Natural Events from Ancient Times to the 19<sup>th</sup> Century (2016–2020)**

Project leader: Darejan Kldiashvili

Project group: Giorgi Bagrationi, Maia Diasamidze, Manuchar Guntsadze, Mikheil Elashvili (Ilia State University), Giorgi Kirkitadze (Ilia State University), Darejan Kldiashvili

The project with the title *Natural Disasters in the South Caucasus: Historical Earthquakes and Associated Natural Events from Ancient Times to the 19<sup>th</sup> Century*, conducted between 2016 and 2020, focused on the seismic history of a highly active region encompassing present-day Georgia, Armenia, Azerbaijan, parts of North Iran, and Eastern Turkey. This region lies within the seismically active Mediterranean zone, where historical earthquakes have had profound impacts, including ecological catastrophes, significant loss of life, political and economic turmoil, and shifts in demographic patterns due to mass migrations.

Georgian historical records, along with archaeological and architectural evidence, offer invaluable insights into past seismic events. The interdisciplinary nature of this project aimed to uncover, document, and analyse these historical earthquakes, combining written sources and material culture to establish detailed information about these events. The project's primary goal

was to contribute to the understanding of seismic hazards in the South Caucasus and its surrounding areas by collecting and analysing historical data on earthquakes from ancient times up to the 19<sup>th</sup> century. Here are the key research directions and achievements of the project:

- Collection of seismic data from written sources: The project gathered seismic data from historical written sources in various languages, identifying specific earthquake events described in these texts. This involved careful examination and interpretation of historical documents, chronicles, and other literary sources to extract reliable seismic information.
- Source analysis, dating, and systematisation: The collected reports were analysed to ensure accurate dating and systematisation of seismic events. This process was crucial in creating a coherent historical timeline of earthquakes in the region, enabling researchers to track the frequency and severity of seismic activity over centuries.
- Identification of seismic risk sites and damaged cultural heritage: Using the collected material, the project identified locations with historical seismic risk and cultural heritage monuments that had been damaged by earthquakes. This aspect of the research highlighted the intersection between natural disasters and the preservation of cultural heritage, emphasising the need for protective measures in vulnerable areas.
- Geographical and seismological analysis: The project aimed to precisely determine the dates, coordinates, and affected areas of historical earthquakes. This included creating macroseismic effect distribution maps, which visually represented the impact zones of significant seismic events. The maps also helped in estimating the epicentres and magnitudes of the historical earthquakes.
- Bilingual geo-information documented database: A comprehensive bilingual (Georgian-English) database was developed, documenting the geo-information related to historical earthquakes in the South Caucasus region. This database serves as a critical resource for researchers and policymakers involved in seismic risk assessment and disaster preparedness.
- Publication of the findings: A detailed monograph was published which compiles the research findings and offers an in-depth analysis of the seismic history of the region. This work contributes to the scholarly understanding of how ancient societies in Georgia and the broader South Caucasus region responded to and were shaped by natural disasters.<sup>5</sup>

This project has made important strides in enhancing the understanding of the seismic history of the South Caucasus, contributing valuable data for seismic hazard assessment and the preservation of cultural heritage. By integrating historical, archaeological, and seismological research, it provides a model for interdisciplinary studies aimed at mitigating the risks associated with natural disasters in regions with rich historical legacies.

---

<sup>5</sup> Kldiashvili 2023.

## **8. Medical and Veterinary Manuscripts in the Depositories of Georgia and Abroad (2017–2022)**

Project Leader: Liana Samkurashvili

Project group: Tamar Abuladze, Luiza Gabunia (Medical University), Irina Gogonaia, Maia Kereselidze (European University), Natia Khizanishvili, Liana Samkurashvili, Lela Shatirishvili.

The project *Medical and Veterinary Manuscripts in the Depositories of Georgia and Abroad*, carried out from 2017 to 2022 in collaboration with Tbilisi State Medical University (Georgia) and European University (Tbilisi, Georgia), sought to systematically catalogue and study the rich body of medical and veterinary knowledge preserved in Georgian manuscripts and collections abroad. This initiative recognised medical knowledge as a fundamental component of cultural history, historically intertwined with philosophy, and reflected across various disciplines including theology, natural sciences, and law. The project represents the first systematic effort to catalogue and create a database of these manuscripts, providing accessible and unified information for specialists in the field. This resource is invaluable for researchers seeking to explore the rich medical tradition preserved in Georgian manuscripts.

The manuscripts explored in this project encompass a wide array of materials, including religious and secular texts, general medical and pharmacological treatises, veterinary works, *karabadins* (pharmacopoeias), educational and family medical books, encyclopedias, references, formulas, and specialised dictionaries. These works are widely represented in Georgian manuscripts and mixed collections, offering a comprehensive view of the evolution of medical knowledge in Georgia. Georgian manuscripts, along with historical documents and legal monuments, thus provide insights into the long history of medical knowledge in the region. They reflect the intellectual and cultural developments of Georgian society, the integration of folk medicine with professional medical practices, and the influence of external cultural and historical trends. Despite the richness of this material, most manuscripts of medical and veterinary content had not been thoroughly studied in the context of the history of medicine prior to this project. The key features and achievements of the project are as follows:

- **Creation of a database and a catalogue:** The project resulted in the first systematic and unified database and catalogue of medical and veterinary manuscripts in Georgia and abroad. This resource is accessible to specialists in the field, providing a comprehensive overview of the manuscripts, their contents, and their historical significance.
- **Historical and cultural contextualisation:** The project placed these manuscripts within the broader context of Georgian intellectual history, highlighting the relationship between medical knowledge and other cultural domains such as philosophy, theology, and law. This approach underscored the interconnectedness of medical practices with the cultural and historical evolution of Georgian society.
- **Publication of the findings:** The results of the project were published in book form.<sup>6</sup>

This project not only brought significant attention to the medical and veterinary manuscripts preserved in Georgian depositories but also advanced the broader field of Digital Humanities

---

<sup>6</sup> Abuladze et al. 2022.



by integrating ancient texts into modern, accessible formats. Through these efforts, the project has made substantial contributions to the preservation and study of Georgia's medical heritage and its cultural legacy.

## 9. Contributions of early-career researchers

An integral aspect of the Digital Humanities work at the National Center of Manuscripts is the involvement of younger members of the Center in the digital publication of ancient texts, particularly focusing on biblical texts like the Bible.

Natia Dundua is a recognised Georgian scholar specialising in Digital Humanities and biblical studies. One of her notable contributions is her work on creating a parallel electronic edition of the Georgian versions of *Ecclesiastes*. This project was part of a broader EU-funded initiative aimed at integrating ancient Georgian biblical texts into digital formats, making them more accessible for scholarly research. In this project, each word in the Georgian versions of *Ecclesiastes* was annotated grammatically, and an English translation was provided, offering a comprehensive tool for researchers studying the text. The platform developed for this purpose was designed to be user-friendly, with a graphical interface that allowed for detailed annotations and automated updates to the database.

Natia Mirotadze is particularly known for her work on ancient Georgian translations of biblical texts. One of her key contributions is her involvement in the *Biblical Synopsis Online* project, which focuses on aligning corresponding segments of different textual traditions of the Bible. In this project, she is responsible for preparing editions of the ancient Georgian translations of the *Book of Esther* and the *First Book of Kings*. Her work in this area is notable for its deep exploration of the textual variations and the historical context of these translations, which often combine elements from different Greek versions of the biblical texts.

Above, we have highlighted several key projects and initiatives in Digital Humanities undertaken by the National Center of Manuscripts. All of the aforementioned digital databases are accessible via the Center's website. However, the website is currently undergoing reconstruction. Once the technical updates are completed, these databases will be freely available to the public.

## References

Abashidze & Surguladze (1991): თ. აბაშიძე, მ[ზია] სურგულაძე, „წინასწარი ანგარიში ისტორიული დოკუმენტების ავტომატიზებული ბანკების შედგენის შესახებ [Preliminary Report on Compilation of Automated Databases of Historical Documents]“, *მრავალთავი [Mravaltavi]* 16, 189–195. <https://iverieli.nplg.gov.ge/handle/1234/259028>.

Abuladze et al. (2022): სამედიცინო და ვეტერინარული ხელნაწერები საქართველოსა და უცხოეთის სიძველეთსაცავებში, თ. აბულაძისა და ლ. სამყურაშვილის რედაქციით, ავტორები: ლიანა სამყურაშვილი, ირინა გოგონაია, თამარ აბულაძე, ლელა შათირიშვილი, ნათია ხიზანიშვილი; ლუიზა გაბუნია, მაია კერესელიძისა და კარენ მულკიჯანიანის მონაწილეობით, თბილისი: კ. კეკელიძის სახელობის საქართველოს ხელნაწერთა ეროვნული ცენტრი / *Medical and Veterinary Manuscripts in Antiquities of Georgia and Abroad*, edited by Tamar Abuladze and Liana Samkurashvili, authors:

Liana Samkurashvili, Irina Gogonaia, Tamar Abuladze, Lela Shatirishvili, Natia Khizanishvili, with the participation of Luiza Gabunia, Maia Kereselidze and Karen Mulkijanian, Tbilisi: K. Kekelidze National Center of Manuscripts of Georgia.

- Aleksidze et al. (2015): თარიღიანი ქართული ხელნაწერები, 9–16 სს. / *Dated Georgian Manuscripts, 9<sup>th</sup>–16<sup>th</sup> cc. An Album*, prepared by Zaza Aleksidze, Ketevan Asatiani, Buba Kudava, Elene Machavariani, Thamar Otkhmezuri, Maia Surguladze, Dali Chitunashvili and Temur Jojua, edited by Zaza Aleksidze and Buba Kudava, translated by Nikoloz Aleksidze (in Georgian and English languages), Tbilisi: K. Kekelidze Georgian National Center of Manuscripts.
- Gagnidze et al. (2015): *Georgian Manuscripts in the Regions of Georgia. Catalogue*. Materials for publication prepared by Grigol Gagnidze, Shalva Gloveli, Darejan Gogashvili, Maia Karanadze, Thamar Otkhmezuri, Nestan Chkhikvadze, editor Shalva Gloveli (in Georgian and English languages), Tbilisi: K. Kekelidze Georgian National Center of Manuscripts.
- Kldiashvili (2011): მონასტრული ისტორიული დოკუმენტები საქართველოს სიბველეთსაცავებიდან [*Illuminated Historical Documents from the Antiquities of Georgia*], დარეჯან კლდიაშვილის რედაქციით, თბილისი: ფაგორიტი პრინტი.
- (2023): დარეჯან კლდიაშვილი, მიწისძვრები ძველ საქართველოში. ქართული საისტორიო წყაროები. IV–XVIII სს., რედაქტორები თამარ აბულაძე და ნოდარ ბახტაძე, თბილისი: კ. კეკელიძის სახელობის საქართველოს ხელნაწერთა ეროვნული ცენტრი / *Earthquakes in ancient Georgia. Georgian historical sources. IV–XVIII centuries*, editors Tamar Abuladze and Nodar Bakhtadze, Tbilisi: K. Kekelidze National Center of Manuscripts of Georgia.

დიგიტალური ჰუმანიტარია და კ. კეკელიძის სახელობის  
საქართველოს ხელნაწერთა ეროვნული ცენტრი  
მაია მაჭავარიანი

(კ. კეკელიძის სახელობის საქართველოს ხელნაწერთა ეროვნული ცენტრი)

DOI: <https://doi.org/10.62235/dk.3.2024.8509>

[matchavarianimn@yahoo.com](mailto:matchavarianimn@yahoo.com) || ORCID: [0009-0008-7155-1068](https://orcid.org/0009-0008-7155-1068)

სტატიაში წარმოდგენილია კ. კეკელიძის სახელობის ხელნაწერთა ინსტიტუტში, შემდგომში – კ. კეკელიძის სახელობის ხელნაწერთა ეროვნულ ცენტრში განხორციელებული, ციფრულ ჰუმანიტარიასთან დაკავშირებული რამდენიმე პროექტი.

**1. ქართული დამწერლობა და პალეოგრაფიული ალბომი (2012–2014 წწ.)**

ხელმძღვანელები: ზაზა ალექსიძე, ბუბა კუდავა

სამუშაო ჯგუფი: ზაზა ალექსიძე, ქეთევან ასათიანი, ბუბა კუდავა, ელენე მაჭავარიანი, თამარ ოთხმეზური, თეიმურაზ ჯოჯუა, დალი ჩიტუნაშვილი

პროექტი მიზნად ისახავდა ქართული თარიღიანი ხელნაწერების პალეოგრაფიული ანალიზის საფუძველზე უთარილო ნუსხების დათარიღების მეთოდოლოგიების შემუშავებას. საქართველოსა და სხვა ქვეყნების სიძველეთსაცავებში მოძიებული მასალა განთავსდა ქართულ ხელნაწერთა პალეოგრაფიულ მონაცემთა ბაზაში, რის შედეგადაც წარმოჩნდა ქართული დამწერლობის განვითარების სხვადასხვა ეტაპი. პროექტის შედეგები ასევე წარმოდგენილია ალბომში – „თარიღიანი ქართული ხელნაწერები“.

**2. ქართული, სპარსული და ოსმალური მოხატული დოკუმენტები საქართველოს სიძველეთსაცავებიდან (მონაცემთა ბაზა, ვებ-გვერდი, ალბომი) (2007–2011 წწ.)**

ხელმძღვანელი: დარეჯან კლდიაშვილი

სამუშაო ჯგუფი: დარეჯან კლდიაშვილი, თამარ აბულაძე, ცისანა აბულაძე, ირინა კოშორიძე (საქართველოს ეროვნული მუზეუმი), ზაზა სხირტლაძე (თბილისის სახელმწიფო უნივერსიტეტი)

პროექტი მიზნად ისახავდა საქართველოს სიძველეთსაცავებში დაცული XI-XIX საუკუნეების მოხატული ქართული და სხვაენოვანი დოკუმენტების ისტორიული და კულტურული მნიშვნელობის შესწავლას, მათი მხატვრული განვითარების ქრონოლოგიური ეტაპების და ტიპოლოგიის დადგენას და ამ დოკუმენტების დიგიტალური ბაზის შექმნას. ბაზაში განთავსდა ინფორმაცია 200-ზე მეტი მოხატული დოკუმენტის შესახებ, მომზადდა ამ დოკუმენტების ციფრული ასლებიც. პროექტის შედეგები ასევე გამოქვეყნდა წიგნში „მოხატული ისტორიული დოკუმენტები საქართველოს სიძველეთსაცავებიდან“.

### **3. ასტრონომიული ხელნაწერები საქართველოში (მონაცემთა ბაზა) (2012–2015 წწ.)**

ხელმძღვანელი: ირაკლი სამსონია (ილიას სახელმწიფო უნივერსიტეტი)

სამუშაო ჯგუფი: ირაკლი სამსონია (ილიას სახელმწიფო უნივერსიტეტი), თამარ აბულაძე, ლიანა სამყურაშვილი, ნესტან ჩხიკვაძე

ხელნაწერთა ეროვნული ცენტრისა და ილიას სახელმწიფო უნივერსიტეტის ერთობლივი პროექტი გულისხმობდა საქართველოს სიძველეთსაცავებში დაცული ქართული, სპარსული და არაბული ასტრონომიული ხელნაწერების კვლევას, სისტემატიზაციასა და კატალოგიზაციას, ვებ-გვერდისა და ინტერაქტიური მონაცემთა ბაზის შექმნას. პროექტის ფარგლებში შესწავლილ იქნა 360 ხელნაწერი, ვებ-გვერდზე, ხელნაწერთა ანოტაციებისა და ვიზუალური მასალის გარდა, ასევე განთავსდა ასტრონომიული ტერმინების ლექსიკონი.

### **4. ისტორიული საბუთების დედნების დიגיტიზაცია და მონაცემთა ბაზა (2012–2014 წწ.)**

ხელმძღვანელი: თეა ქართველიშვილი

სამუშაო ჯგუფი: თეა ქართველიშვილი, ნიკოლოზ ჟღენტი, თამაზ გოგოლაძე, ირაკი გელაშვილი, მაია შაორშაძე, ნესტან ბაგაური, ოთარ ჟიჟიაშვილი

პროექტი მიზნად ისახავდა ხელნაწერთა ეროვნულ ცენტრში დაცული ისტორიული საბუთების ანალიზს და ხელმისაწვდომობის უზრუნველყოფას. პროექტის ფარგლებში დამუშავდა და გაციფრულდა 5000 ისტორიული დოკუმენტი. შეიქმნა მონაცემთა ბაზა, სადაც შეტანილია თითოეული დოკუმენტის დაწვრილებითი არქეოგრაფიული აღწერა და შესაძლებელია მონაცემთა მოძიება სხვადასხვა კრიტერიუმის მიხედვით (თარიღი, გეოგრაფიული არეალი და სხვა).

### **5. საქართველოს რეგიონებში დაცული ქართული ხელნაწერები: მონაცემთა ერთიანი ელექტრონული ბაზა (ქართულ და ინგლისურ ენებზე) (2013–2015 წწ.)**

ხელმძღვანელი: შალვა გლოველი

სამუშაო ჯგუფი: ნესტან ჩხიკვაძე, გრიგოლ გაგნიძე (საქართველოს ეროვნული მუზეუმი), შალვა გლოველი, დარეჯან გოგაშვილი, მაია კარანაძე, თამარ ოთხმეხური

საქართველოს ეროვნულ მუზეუმთან თანამშრომლობით განხორციელებული ეს პროექტი მიზნად ისახავდა საქართველოს სხვადასხვა რეგიონის სიძველეთსაცავებში დაცული ხელნაწერების ერთიანი, ორენოვანი (ქართულ-ინგლისური) ელექტრონული კატალოგის და მონაცემთა ბაზის შექმნას, ასევე რეგიონალური ხელნაწერების გასული საუკუნის 60-70-იან წლებში შექმნილი გამოუქვეყნებელი აღწერილობების გადახედვასა და განახლებას. პროექტის შედეგები გამოქვეყნდა ორენოვან კატალოგში, სადაც წარმოდგენილია ინფორმაცია ქვეყნის რეგიონებში დაცული 1305 ქართული ხელნაწერის შესახებ.

**6. ქართული ილუსტრირებული ხელნაწერების გაფორმებაში ჩართული მინიატურების მონაცემთა ბაზა (XI–XIX სს.) (2014–2016 წწ.)**

ხელმძღვანელი: ქეთევან ტატიშვილი.

სამუშაო ჯგუფი: ქეთევან ტატიშვილი, ელენე მაჭავარიანი, ნინო ციციშვილი.

პროექტი გულისხმობდა მოხატული ქართული ხელნაწერების სხვადასხვა მონაცემის კატალოგიზაციასა და ამ ხელნაწერების ხელმისაწვდომობის უზრუნველყოფას. პროექტის ფარგლებში შეიქმნა ორენოვანი (ქართულ-ინგლისური) მონაცემთა ბაზა, რომელიც მეცნიერებს სთავაზობს დაწვრილებით ინფორმაციას როგორც ხელნაწერების, ასევე მათში შემავალი ცალკეული მინიატურების შესახებ. ბაზაში შეტანილია ინფორმაცია როგორც ხელნაწერთა ეროვნული ცენტრში, ასევე საქართველოს სხვა სიძველეთსაცავებში დაცული ხელნაწერების შესახებ.

**7. ბუნებრივი კატასტროფები სამხრეთ კავკასიაში: ისტორიული მიწისძვრები და თანმდევი ბუნებრივი მოვლენები უძველესი დროიდან XIX საუკუნემდე (2016–2020 წწ.)**

ხელმძღვანელი: დარეჯან კლდიაშვილი

სამუშაო ჯგუფი: დარეჯან კლდიაშვილი, მაია დიასამიძე, მანუჩარ გუნცაძე, მისეილ ელაშვილი (ილიას სახელმწიფო უნივერსიტეტი), გიორგი კირკიტაძე (ილიას სახელმწიფო უნივერსიტეტი), გიორგი ბაგრატიონი

პროექტის მიზანი იყო სამხრეთ კავკასიასა და მის მიმდებარე რეგიონებში ისტორიული მიწისძვრების (უძველესი დროიდან XVIII საუკუნის ჩათვლით) შესახებ ინფორმაციის მოძიება – წერილობითი და მატერიალური კვალის ამსახველი მასალის გამოვლენა, თავმოყრა და გაანალიზება. მოპოვებული დამუშავებული მასალა განთავსდა ისტორიული მიწისძვრების მონაცემთა ელექტრონულ ბაზაში, ასევე მომზადდა სამხრეთ კავკასიის რეგიონის ისტორიული მიწისძვრების ერთიანი კატალოგი.

**8. სამედიცინო და ვეტერინარული ხელნაწერები საქართველოსა და საზღვარგარეთის საცავებში (2017–2022 წწ.)**

ხელმძღვანელი: ლიანა სამყურაშვილი.

სამუშაო ჯგუფი: თამარ აბულაძე, ლუიზა გაბუნია (სამედიცინო უნივერსიტეტი), ირინა გოგონაია, მაია კერესელიძე (ევროპის უნივერსიტეტი), ნათია ხიზანიშვილი, ლიანა სამყურაშვილი, ლელა შათირიშვილი

პროექტი, რომელიც განხორციელდა თბილისის სახელმწიფო სამედიცინო უნივერსიტეტსა და ევროპის უნივერსიტეტთან თანამშრომლობით, მიზნად ისახავდა ქართულ ხელნაწერებში დაცული მდიდარი სამედიცინო და ვეტერინარული ცოდნის კატალოგიზაციასა და მონაცემთა ბაზის შექმნას, რომლის საშუალებითაც ეს ინფორმაცია ხელმისაწვდომი იქნებოდა დარგის სპეციალისტებისათვის. პროექტის შედეგები ასევე გამოქვეყნდა ორენოვან გამოცემაში „სამედიცინო და ვეტერინარული ხელნაწერები საქართველოსა და უცხოეთის სიძველეთსაცავებში“.



## 9. ხელნაწერთა ეროვნულ ცენტრის ახალგაზრდა მეცნიერები

ხელნაწერთა ეროვნულ ცენტრის ახალგაზრდა თანამშრომლები ნათია დუნდუა და ნათია მიროტაძე, ასევე ჩართული არიან უძველესი ტექსტების ციფრული გამოცემის საერთაშორისო პროექტებში, მათ შორის ყველაზე მნიშვნელოვანია ბიბლიური ტექსტების ციფრული გამოცემა.

სტატიაში წარმოდგენილია ხელნაწერთა ეროვნული ცენტრის მიერ განხორციელებული მხოლოდ რამდენიმე ძირითადი პროექტი. ყველა ზემოაღნიშნული ციფრულ მონაცემთა ბაზა ხელმისაწვდომია ცენტრის ვებსაიტზე. თუმცა, ვებსაიტი ამჟამად რეკონსტრუქციას გადის. ტექნიკური განახლებების დასრულების შემდეგ, ეს მონაცემთა ბაზები კვლავ იქნება ხელმისაწვდომი ფართო საზოგადოებისთვის.

# Building a Dialectological Lexicological Database of Georgian Cognates for Digital Analysis<sup>1</sup>

Hélène Gérardin (Paris)

DOI: <https://doi.org/10.62235/dk.3.2024.8511>

[helene.gerardin@inalco.fr](mailto:helene.gerardin@inalco.fr) || ORCID: [0009-0003-7406-3901](https://orcid.org/0009-0003-7406-3901)

**Abstract:** This paper examines the challenges involved in creating a lexical database for Georgian dialects. It begins by outlining the methodological approaches to data collection and formatting, followed by an overview of the current version of the database, and its applications for linguistic analysis. Designed to facilitate a wide range of research, the database supports investigations such as Levenshtein distance calculations and diachronic and diatopic comparisons. The long-term goal of this project is to develop an open, accessible resource that can be gradually enriched with new data, advancing computational tools and deepening our understanding of Georgian and Kartvelian dialects.

**Keywords:** Georgian, Dialectology, Levenshtein distances, Diatopic and diachronic comparison

## Introduction

The aim of this paper is to present a Digital Lexical Database of Georgian dialects that I have been building for a few years. The idea was initiated during the IDEX project Linguistic Dynamics of the Caucasus (LaDyCa),<sup>2</sup> which was conducted in 2017–2018 in collaboration of Sorbonne University (Paris) and Ilia State University (Tbilisi).<sup>3</sup> Throughout this project, I was in charge of collecting Georgian data (recording texts and lexical materials) with the intention of managing them via computer programs (e.g. Markov Models, Levenshtein distances). It was within this framework that the idea of focusing on lexical data emerged, leading to the decision to continue the processing to create a comprehensive database for broader purposes. The present article aims to describe the stages of this new task, along with its methodology, challenges and caveats.

After offering a brief overview of Georgian dialectology, I will delve into the issues, methods and challenges of my work. Following that, I will show the results of the created database, starting with the application of Levenshtein Distances to the database – a collaborative endeavour with colleagues from the LaDyCa project. In addition, I will present the current state of the materials and provide some examples of potential uses for the database. In the conclusion, I will attempt to delineate further steps necessary for advancing this research.

---

<sup>1</sup> I am grateful to Jost Gippert and Manana Tandaschwili for their interest in this work and the opportunity they provided me to present it at the conference “Digital Caucasiology – A Change of Paradigm” and subsequently publish it in this journal. I would also like to extend special thanks to Donald Rayfield and George B. Hewitt for their valuable feedback and encouragement.

<sup>2</sup> The IDEX projects (“Excellence Initiatives”, in French “Initiatives d’excellence”) are part of the “Investments for the Future” programs, which are initiatives set up by the State of France and aimed at creating multidisciplinary higher education and research institutions in the country that would be globally competitive. For details as to LaDyCa see Léonard 2019b.

<sup>3</sup> I thank Tamar Makharoblidze and Jean-Léo Léonard for giving me the opportunity to lay the first steps of the work presented here.

## 1. Overview of Georgian dialectology

Georgian exhibits a wide linguistic and geographical variation, resulting in the development of approximately 15–20 dialectal or subdialectal varieties. Several linguists have proposed different classifications,<sup>4</sup> but none of them has achieved unanimous acceptance. According to Jorbenadze 1989, Georgian dialects can be divided into two branches, along an East-West axis, which also corresponds to a historical and geographical boundary:

- Eastern Dialects
  - Mountain dialects: Pshavian, Khevsurian, Tushetian, Mokhevian, Mtiuletian-Gudamaqrian
  - Dialects of the plain: Kartlian, Kakhetian, Ingiloan, Fereydani
  - Meskhian, Javakhian
- Western Dialects
  - Ratchian
  - Imeretian, Letchkhumian
  - Gurian, Adjarian (+ Imerkhevian)

Standard Georgian is based on the Kartlian dialect, which is spoken in the area encompassing both the historical capital (Mtskheta) and the current capital (Tbilisi). Mutual intelligibility between Standard Georgian and the dialects is almost complete. All dialects are oral varieties with no writing tradition. They are seriously endangered, yet they offer data of primary importance for understanding the history and development of Georgian. As such, they constitute an important repository of grammatical and lexical categories that are not or no longer attested in Georgian. For this reason, their description, starting with a review of the lexicon, is urgent and promising.

## 2. Why a lexical dialectological database?

### 2.1. Aims

The current research aims to create a lexical database of cognates including items for all dialectal varieties, and later archiving the database online, facilitating its future expansion. Such work can be useful not only for archiving and processing dialectological materials but also for studying the linguistic variation of Georgian, as well as refining the classification of Georgian dialects. Furthermore, it is an efficient way of exploring and reconstructing the diachrony of the language and enhancing our understanding of the Kartvelian family.

This unique database is thus intended to be used for various subsequent studies and research frameworks.

---

<sup>4</sup> See for instance Chikobava 1952, Shanidze 1957, Dzidziguri 1970 and Jorbenadze 1989; for a cartographic representation and discussions about migrations, see Beridze et al. 2018.

## 2.2. Challenges

The heart of the work involves bringing together lexical units common for all Georgian dialects. This has two methodological implications: first, the need to eliminate borrowings (which are numerous, especially in dialects that are in contact with other languages); and secondly, the identification of words common enough to be encountered in corpora or elicited. Last but not least, it is important to compile a list extensive enough to be considered representative (with no less than a hundred items).

Due to these difficulties and the scarcity of dialectological resources, we must combine a diverse array of sources and compile both written and oral corpora.

## 2.3. The corpus

### 2.3.1 Written sources

Written sources primarily include dictionaries and word lists, usually found appended to text collections and descriptive works. A notable problem of such lists is their tendency to contain lexemes which deviate the most from the standard while common lexemes are often lacking.

Furthermore, collecting the items can also be carried out on the basis of published texts and online corpora.<sup>5</sup> A caveat when using texts is that lexemes used in context usually appear in modified grammatical forms. This requires determining the citation form.

Another common problem across all types of written sources is that transcription systems are not homogeneous (for instance, the sound [w] is transcribed as either *w/ჲ* or *v/ვ* or *u/უ*). This requires adapting the phonetic notation of certain items.

### 2.3.2. Oral sources

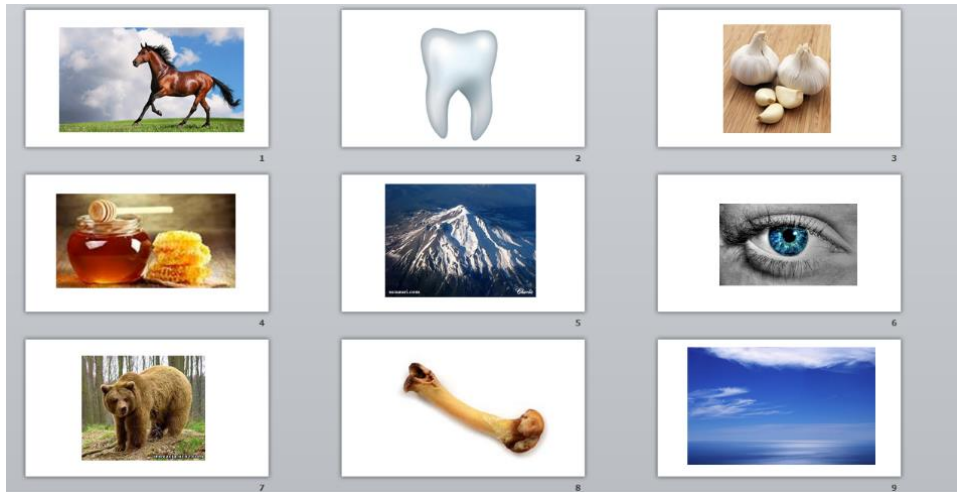
Oral corpus collection involves recording isolated words from speakers in the villages where the dialects are spoken. However, this task is far from being simple. First of all, the number of dialect speakers is dwindling and some areas remain difficult to access. Secondly, in the territory of Georgia, almost all dialect speakers are able (and used) to speak Standard Georgian, due to its spread through secondary education, the media, new communication tools, and recent migrations. As a result, speakers may not always be aware of the boundary between dialect and standard. However, dialects are still often associated in people's minds with inferior and non-prestigious ways of speaking. A consequence of this is that even individuals who only speak in a dialect automatically switch to standard Georgian as soon as they are recorded. Combining dialect and recording material is sometimes a real challenge for the linguist!

For all these reasons, in order to minimise interference with Standard Georgian, I had to develop an alternative elicitation protocol. I created a slideshow of pictures representing the target words, so that the speakers only had to mention what they saw in the pictures. Each slideshow comprises approximately 40 items, so that each inquiry lasts approximately 20–40 minutes.

---

<sup>5</sup> Electronic corpora are easily available, see for instance the Georgian Dialect Corpus (<http://corpora.co>) and the Georgian National Corpus (<http://gnc.gov.ge>), not forgetting the large database of TITUS (<https://titus.uni-frankfurt.de/texte/texte2.htm#georgant>).

Keeping word lists concise is essential, since speakers tend to get tired after 40 words within the session. The pictures should be easily identifiable, and the speakers are invited to label the items and avoid discussing them as much as possible in order to reduce interference from Standard Georgian. The screenshot in Fig. 1 presents an extract from the slideshow:



**Fig. 1: Sample from the slideshow presenting lexical items**

Such lexical elicitations can still pose a number of challenges. For instance, some lexemes cannot be conveyed through images, which requires seeking alternative methods.

I have tried to devise other ways for the speakers to produce the words as spontaneously as possible. While visual representations work well for numbers and colors, they prove difficult for qualifying adjectives, verbs or some nouns conveying functions or relations when not associated with any context. In this case, I had to communicate with the speakers in Standard Georgian to explain what the target term was. This gave rise to the use of various strategies, such as prompting people to guess words through completing sentences, engaging in logical enumerations, or even translating items from Russian or English. In any case, the experience gained from fieldwork in dialectology underlines the importance of avoiding pronouncing directly the Standard Georgian word because on hearing such forms, most speakers will automatically repeat them, making it impossible to capture the ‘true’ dialectal form. The screenshot in Fig. 2 presents a selection of such adapted slides taken from the end of the slideshow.



**Fig. 2: Sample of adapted items**



### 3. First results

#### 3.1 The current version of the Database

Once collected, the materials must be transcribed to fill the database. The result is a table listing all the items classified by dialect and accompanied by their translation into English and French. The items are represented by their Standard Georgian lemma and organised in alphabetic order. To facilitate comparison and diachronic analysis, we have also added the forms in Old Georgian as well as reconstructed Kartvelian etymons when they are known (at the left). In addition, in order to broaden the base to include other Kartvelian languages, columns were created for Megrelian, Laz and Svan (positioned to the right of the Georgian dialects). The lexical Database, after formatting, exhibits the structure shown in Fig. 3 (extract).

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	AA	AB	AC	AD	AE					
1	Etymo	Gr	Old Gr	Khev	Psh.	Tian.	Mokl	Mtšul.	Gud	Tush.	Kartl.	Kakh.	Kiziq.	Fer	Ing.	Djav.	Samtsk	Ratch.	Im	Lech.	Adj.	Tao	Imerkh	Gur.	Mingr.	Laz	Svan	Balze	Lachk	Com	Meanin					
2	*kac	kaci	kaci	kaci	kaci	kaci	kaci	kaci	kaci	kaci	kaci	kaci	kaci	kaci	kaci	kaci	kaci	kaci	kaci	kaci	kaci	kaci	kaci	kaci	kaci	koči	koči	čāši.	čāš			Svan	man			
3	*kakali	kakali									kakali	kakali	kakali	kakali	kakali	kakali	kakali	kakali	kakali	kakali	kakali	kakali	kakali	kakali	kakali	kakali	kakali	kakali	kakali	kakali	kakali	kakali	kakali	Mingr	nut	
4		kali	kali	kali	kali	kali	kali	kali	kali	kali	kali	kali	kali	kali	kali	kali	kali	kali	kali	kali	kali	kali	kali	kali	kali	kali	kali	kali	kali	kali	kali	kali	kali	Khev	women	
5		karga	kargad	kargə	karga(d)	karga	karga(t)				karka	karga	karkat	[nama]	karga	kargat	karkat	karkat	karkat	karkat	karkat	karkat	karkat	karkat	karkat	karkat	karkat	karkat	karkat	karkat	karkat	karkat	karkat	karkat	Ingil.	wind
6		kari	kari	kari	kari	kari	kari	kari	kari	kari	kari	kari	kari	kari	kari	kari	kari	kari	kari	kari	kari	kari	kari	kari	kari	kari	kari	kari	kari	kari	kari	kari	kari	Ingil.	wind	
7		kargi	kargi	kargi	kargi	kai	kargi	karki	kai	karga	karki	kai,	k:	kai,	k:	[nan	kaj	kai,	k:	kai	karki,	k:	kaj,	karki	kaj	kaj	karka								good	
8	*katan	katari	katami		katami	katan	katami				katam	katam	kata	katam	katam	katami	katami	katami	katami	katami	katami	katami	katami	katami	katami	kotom	kotom	katal					chicken			
9		keipi	keipi	kaibi	keipi	keipi	keipi	keipi	keipi	keipi	keipi	keipi	keipi	keipi	keipi	keipi	keipi	keipi	keipi	keipi	keipi	keipi	keipi	keipi	keipi	keipi	keipi	keipi	keipi	keipi	keipi	keipi	keipi	feast		
10		kidev	kidev	kide	kide	kider	kiden	kidav,	kida	kiden	kidev	kiden	kide,	kide	kide	kide	kide,	kide	kide,	kide	kido,	kide	kide	kide	kido	kide	kido	kide	kido	kide	kido	kide	kido	again		
11	*kjde	kilde	klde	kilde	klde	kilde	klde			kdej.	kilde	klde	klde	[daama	klde	klde	klde,	klde	klde,	klde	klde,	klde	klde	klde	klde	klde	kyrde,	kyrde	kyrde	kyrde,	kyrde	kyrde	kyrde	rock		
12	*kmar	kmari	kmari	kmari	kmari	kmari	kmari	kmari	kmari	kmari	kmari	kmari	kmari	kmari	kmari	kmari	kmari	kmari	kmari	kmari	kmari	kmari	kmari	kmari	kmari	kmari	kmari	kmari	kmari	kmari	kmari	kmari	kmari	kmari	husband	
13		kvali	kvali	kvali	kvali	koli	koli	koli	kvali		kvali	kvali	kvali	[rade	pa]	kvali	kvali	kvali	kvali	kvali	kvali	kvali	kvali	kvali	kvali	kvali	kvali	kvali	kvali	kvali	kvali	kvali	kvali	trace		
14		kveqana	koqə	qveqana.	koqə	koqan	koqə	qveqana	koqə	qveqana	qveqə	qveqə	[keš	[dnt	koqan	koqana	kveqana	koqana	koqana	koqana	koqana	koqana	koqana	koqana	koqana	koqana	koqana	koqana	koqana	koqana	koqana	koqana	koqana	koqana	< *kv	world,
15		korčili	qorči	korčili	qorči	korčili	korčili	korčili	korčili	korčili	korčili	korčili	korčili	korčili	korčili	korčili	korčili	korčili	korčili	korčili	korčili	korčili	korčili	korčili	korčili	korčili	korčili	korčili	korčili	korčili	korčili	korčili	korčili	korčili	marriage	
16	*kowz	kovzi	kozi	kovzi	kovzi	kozi	kovzi	kovzi	kozi	kovzi	kovzi	kovzi	kovzi	kovzi	kovzi	kovzi	kovzi	kovzi	kovzi	kovzi	kovzi	kovzi	kovzi	kovzi	kovzi	kovzi	kovzi	kovzi	kovzi	kovzi	kovzi	kovzi	kovzi	spoon		
17	*kva	kva	kva	kva	kva	kva	kva	kva	kva	kva	kva	kva	kva	kva	kva	kva	kva	kva	kva	kva	kva	kva	kva	kva	kva	kva	kva	kva	kva	kva	kva	kva	kva	rock		
18	*kverc	kvercx	kvercx	kvercx	kvercx	kvercx	kvercx	kvercx	kvercx	kvercx	kvercx	kvercx	kvercx	kvercx	kvercx	kvercx	kvercx	kvercx	kvercx	kvercx	kvercx	kvercx	kvercx	kvercx	kvercx	kvercx	kvercx	kvercx	kvercx	kvercx	kvercx	kvercx	kvercx	Kartv	egg	
19	*kbil	kbili	kbili	kbili	kbili	kbili	kbili	kbili	kbili	kbili	kbili	kbili	kbili	kbili	kbili	kbili	kbili	kbili	kbili	kbili	kbili	kbili	kbili	kbili	kbili	kbili	kbili	kbili	kbili	kbili	kbili	kbili	kbili	tooth		
20		kvira	kvira	kvira	kvira	kvira	kvira	kvira	kvira	kvira	kvira	kvira	kvira	kvira	kvira	kvira	kvira	kvira	kvira	kvira	kvira	kvira	kvira	kvira	kvira	kvira	kvira	kvira	kvira	kvira	kvira	kvira	kvira	kvira	Khev	week
21		lamazi	lama	lama	lama	lama	lama	lama	lama	lama	lamazi	lamazi	lamazi	lamazi	lamazi	lamazi	lamazi	lamazi	lamazi	lamazi	lamazi	lamazi	lamazi	lamazi	lamazi	lamazi	lamazi	lamazi	lamazi	lamazi	lamazi	lamazi	lamazi	lamazi	Mokh	beautif
22		loqa	[kuri]	laqva	laqva,	laqva,	laqo	laqo	laqo	laqo	laqo	laqo	laqo	laqo	laqo	laqo	laqo	laqo	laqo	laqo	laqo	laqo	laqo	laqo	laqo	laqo	laqo	laqo	laqo	laqo	laqo	laqo	laqo	but		
23		magr	magram	magr	magram	magr	magra(m)	magr	magram	magr	magr	magr	magr	magr	magr	magr	magr	magr	magr	magr	magr	magr	magr	magr	magr	magr	magr	magr	magr	magr	magr	magr	magr	magr	check	
24	*mam	mami	mamaj	mam	mama	mami	mama	mama	mami	mama	mami	mama	mami	mama	mami	mama	mami	mama	mami	mama	mami	mama	mami	mama	mami	mama	mami	mama	mami	mama	mami	mama	mami	Mingr	father	
25		0	13	2	4	17	2	0	12	9	15	1	2	3	8	8	4	3	6	13	11	15	6	15	14	15	18	23	23	14	0	0				

Fig. 3: Current database after formatting

The current version of the Database includes 243 items, representing approximately 5,000–6,000 items. Some fields still require completion.

#### 3.2 Application of Levenshtein Distances

The first instance of putting this database to practical use involved the application of Levenshtein Distances. The procedure was conducted during the LaDyCa Project by Laure Picard, Jean-Léo Léonard and myself, using the Gephi software.<sup>6</sup> The results of this study were presented during the project<sup>7</sup> and subsequently partially published.<sup>8</sup>

According to Beijering, Gooskens & Heeringa (2008), the Levenshtein algorithm is a string-based distance measure that quantifies the differences between the (phonetical) shapes of corresponding words in different dialects or closely related languages. It calculates the minimal costs required to change a string of segments into another by means of insertions, deletions or substitutions. The resulting analysis of the Georgian data by Laure Picard is shown in the diagram in Fig. 4.

<sup>6</sup> The *Open Graph Viz Platform*, see <https://gephi.org>.

<sup>7</sup> Picard, Gérardin & Léonard 2018.

<sup>8</sup> Léonard 2019 and Léonard & Makharoblidze 2022.

	Grg	Khev.	Pch.	Mokh.	Mtiul.	Touche	Kartl.	Kakh.	Kiziq.	Ferey.	Ingil.	Meskh.	Ratch.	H-Iméret	B-Iméret	Letchkh.	Adj.	Tao	Imerkh.	H.-Gour.	B.-Gour.
Grg		0.123	0.093	0.15	0.122	0.305	0.127	0.079	0.093	0.183	0.518	0.104	0.119	0.162	0.19	0.25	0.225	0.239	0.176	0.264	0.255
Khev.	0.123		0.138	0.187	0.164	0.26	0.198	0.168	0.142	0.192	0.49	0.183	0.201	0.224	0.247	0.262	0.274	0.234	0.208	0.324	0.309
Pch.	0.093	0.138		0.147	0.116	0.283	0.144	0.105	0.088	0.178	0.478	0.137	0.151	0.177	0.212	0.261	0.241	0.257	0.165	0.312	0.304
Mokh.	0.15	0.187	0.147		0.109	0.35	0.196	0.174	0.164	0.172	0.442	0.198	0.223	0.253	0.25	0.313	0.318	0.282	0.216	0.363	0.343
Mtiul.	0.122	0.164	0.116	0.109		0.337	0.176	0.128	0.118	0.183	0.458	0.169	0.188	0.211	0.232	0.298	0.273	0.266	0.181	0.301	0.305
Touche	0.305	0.26	0.283	0.35	0.337		0.336	0.336	0.305	0.337	0.478	0.334	0.34	0.337	0.353	0.384	0.372	0.399	0.32	0.404	0.391
Kartl.	0.127	0.198	0.144	0.196	0.176	0.336		0.097	0.118	0.172	0.461	0.134	0.143	0.172	0.196	0.235	0.242	0.236	0.164	0.245	0.248
Kakh.	0.079	0.168	0.105	0.174	0.128	0.336	0.097		0.036	0.19	0.522	0.114	0.143	0.191	0.205	0.252	0.208	0.224	0.163	0.277	0.271
Kiziq.	0.093	0.142	0.088	0.164	0.118	0.305	0.118	0.036		0.172	0.504	0.14	0.16	0.205	0.213	0.259	0.226	0.207	0.143	0.291	0.283
Ferey.	0.183	0.192	0.178	0.172	0.183	0.337	0.172	0.19	0.172		0.436	0.181	0.202	0.237	0.254	0.339	0.265	0.265	0.224	0.353	0.337
Ingil.	0.518	0.49	0.478	0.442	0.458	0.478	0.461	0.522	0.504	0.436		0.495	0.511	0.507	0.504	0.566	0.513	0.519	0.528	0.545	0.537
Meskh.	0.104	0.183	0.137	0.198	0.169	0.334	0.134	0.114	0.14	0.181	0.495		0.137	0.173	0.202	0.276	0.261	0.289	0.215	0.29	0.288
Ratch.	0.119	0.201	0.151	0.223	0.188	0.34	0.143	0.143	0.16	0.202	0.511	0.137		0.144	0.168	0.19	0.221	0.273	0.204	0.234	0.241
H-Iméret	0.162	0.224	0.177	0.253	0.211	0.337	0.172	0.191	0.205	0.237	0.507	0.173	0.144		0.125	0.219	0.257	0.275	0.222	0.229	0.235
B-Iméret	0.19	0.247	0.212	0.25	0.232	0.353	0.196	0.205	0.213	0.254	0.504	0.202	0.168	0.125		0.175	0.249	0.288	0.217	0.2	0.192
Letchkh.	0.25	0.262	0.261	0.313	0.298	0.384	0.235	0.252	0.259	0.339	0.566	0.276	0.19	0.219	0.175		0.243	0.357	0.255	0.245	0.26
Adj.	0.225	0.274	0.241	0.318	0.273	0.372	0.242	0.208	0.226	0.265	0.513	0.261	0.221	0.257	0.249	0.243		0.23	0.226	0.288	0.284
Tao	0.239	0.234	0.257	0.282	0.266	0.399	0.236	0.224	0.207	0.265	0.519	0.289	0.273	0.275	0.288	0.357	0.23		0.182	0.301	0.323
Imerkh.	0.176	0.208	0.165	0.216	0.181	0.32	0.164	0.163	0.143	0.224	0.528	0.215	0.204	0.222	0.217	0.255	0.226	0.182		0.276	0.27
H.-Gour.	0.264	0.324	0.312	0.363	0.301	0.404	0.245	0.277	0.291	0.353	0.545	0.29	0.234	0.229	0.2	0.245	0.288	0.301	0.276		0.101
B.-Gour.	0.255	0.309	0.304	0.343	0.305	0.391	0.248	0.271	0.283	0.337	0.537	0.288	0.241	0.235	0.192	0.26	0.294	0.323	0.27	0.101	

Fig. 4: Results following the application of Levenshtein’s algorithm

It may be striking that most numbers are very close to zero. This means that Georgian dialects exhibit a closer proximity to one another than is typically observed among dialects in other languages. A further analysis shows that the differences between the coefficients corroborate the conclusions of the Georgian dialectologists: there is a clear split between Eastern and Western dialects, as was convincingly argued by Jorbenadze (1989). The diagram in Fig. 5 shows the corresponding hierarchical clustering dendrogram.

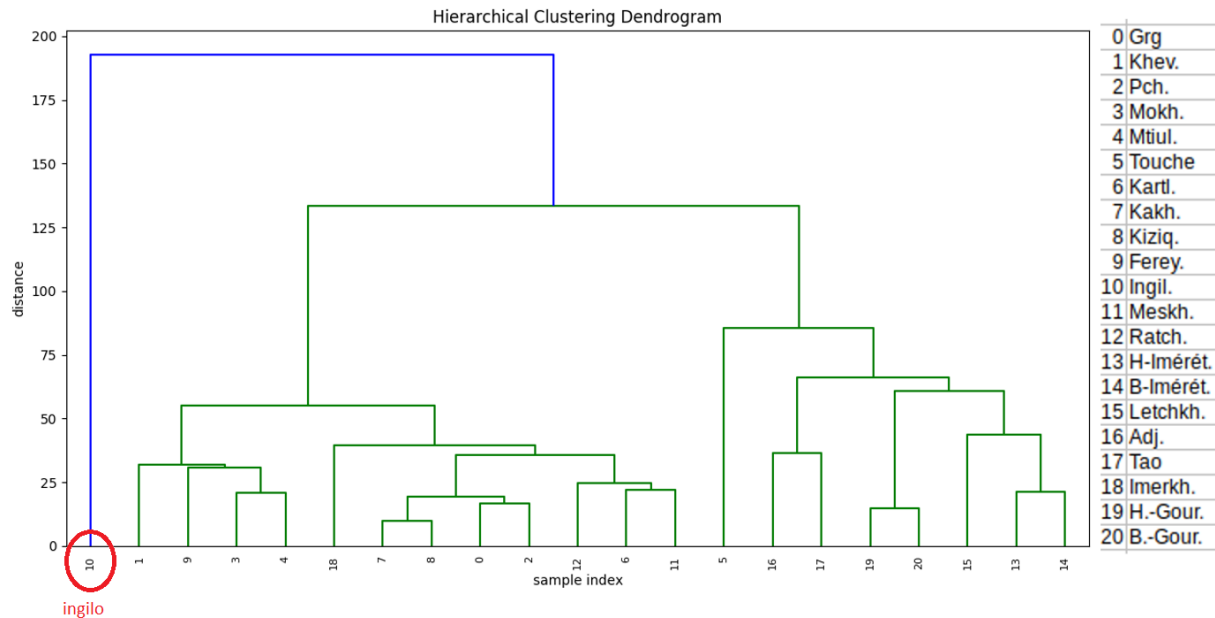


Fig. 5: Results with hierarchical clustering dendrogram

These conclusions can also be presented differently, namely by applying ponderation criteria, as shown in Fig. 6.

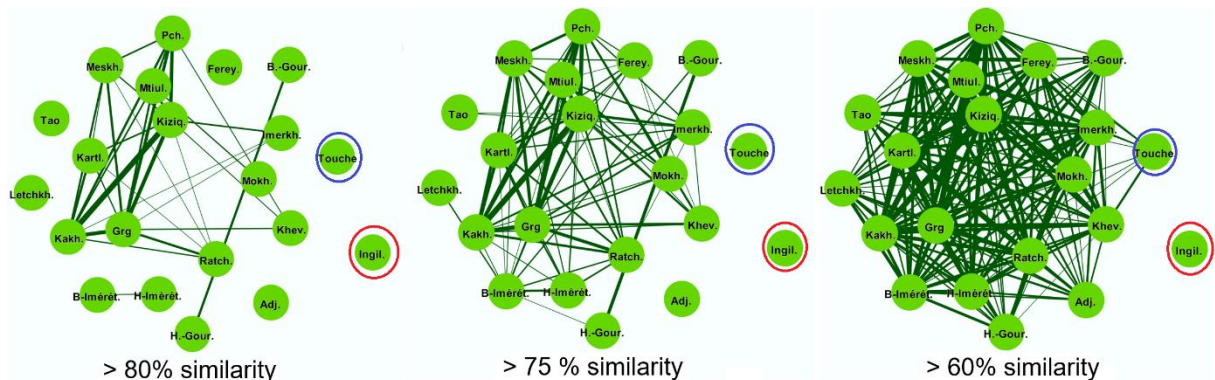


Fig. 6: Results after application of ponderation criteria



Furthermore, with regard to fieldwork, elicitations are more manageable than for Georgian dialects, as the units differ more, and the risk of interference with Standard Georgian is smaller.

Once a significant number of items have been gathered (without waiting for the database to be exhaustively filled), the next step is to archive the data online. Here, a digital approach is highly suitable because the large number of cells (especially columns, due to the large number of dialect varieties) makes paper printing practically impossible. Another advantage of hosting the database online would be the possibility to open it for continual enriching over time. In any case, the archive must include a reference section listing all the sources used for each variety.

## 5. Conclusion

The current paper presents work in progress focused on creating a lexical database of Georgian cognates that I intend to make available to researchers with the aim of promoting the integration of dialectal data in Kartvelological studies. After discussing the main issues and methodological aspects of data collecting and formatting, I have provided several examples of application, among them of Levenshtein Distances. Other potential uses could entail using other distance algorithms or pursuing diachronic comparison. In the future, my aim is to make this base as comprehensive as possible and archive it online. The long-term goal is to create an open database to be gradually expanded.

## References

- Beijering, Gooskens & Heeringa (2008): Karin B., Charlotte G., Wilbert H., “Predicting intelligibility and perceived linguistic distance by means of the Levenshtein algorithm”, *Linguistics in the Netherlands* 25, 13–24.
- Beridze & Nadaraia (2009): Marine B., David N., “The corpus of Georgian dialects”, in *Proceedings of the NLP, Corpus Linguistics, Corpus Based Grammar Research. Fifth International Conference*, Bratislava: Tribun, 25–35.
- Beridze et al. (2018): Marine B., Zakharia Pourtskhvanidze, Lia Bakuradze, David Nadaraia, “Interactive Visualization of Dialectal Lexis Perspective of Research Using the Example of Georgian Electronic Dialect Atlas”, in *Proceedings of the XVIII EURALEX International Congress*, Ljubljana: Ljubljana University Press, Faculty of Arts, 931–939.
- Chikobava (1952): არნოლდ ჩიქობავა, ენათმეცნიერების შესავალი [Introduction to Linguistics], Tbilisi: Tbilisi University Press. <https://iverieli.nplg.gov.ge/handle/1234/411129>.
- Dzidziguri (1970): შოთა ძიძიგური, ქართული დიალექტოლოგიური ძიებანი [Georgian Dialectological Research]. Tbilisi: Ganatleba.
- Gigineishvili, Topuria & Kavtaradze (1961): ივანე გიგინეიშვილი, ვარლამ თოჭურია, ივანე ქავთარაძე, ქართული დიალექტოლოგია [Georgian Dialectology]. Tbilisi: Tbilisi University Press. <https://iverieli.nplg.gov.ge/handle/1234/506997>.
- Jorbenadze (1989): ბესარიონ ჯორბენაძე, ქართული დიალექტოლოგია [Georgian Dialectology], I. Tbilisi: Mecniereba.
- Léonard (2019a): Jean-Léo L., “Méthodes pour l’analyse et la documentation des langues et dialectes kartvèles, à la lumière de la Théorie des Dynamiques Langagières / Théorie de la Complexité [Methods for the Analysis and Documentation of Kartvelian Languages and Dialects, in Light

- of the Theory of Language Dynamics / Complexity Theory]”, in *Proceedings of the Fifth International Scientific Conference Language and Culture*. Kutaisi: Akaki Tsereteli State University, 625–650.
- (2019b): “Le projet LaDyCa (Language Dynamics in the Caucasus, IDEX Emergence 2017–18): avancées et résultats”, *Kadmos* 9, 258–284. <https://doi.org/10.32859/kadmos/9/258-284>.
- Léonard & Makharoblidze (2022): Jean-Léo L. & Tamar M., “Hints at Georgian Dialect History: A Study in Miniature”, in: Tamar Makharoblidze (ed.), *Issues in Kartvelian Studies*, Wilmington: Vernon Press, 3–28.
- Mart'irosovi (1985): არამ მარტიროსოვი, “ქართული დიალექტური ლექსიკის შესწავლისა და ლექსიკონების შედგენის ძირითადი საკითხები” [Key Issues in Studying Georgian Dialect Vocabulary and Compiling Dictionaries]. *Ibero-Caucasian Linguistics* 23, 139–148.
- Picard, Gérardin & Léonard (2018): Flore P., Hélène G., Jean-Léo L., “Levenshtein Algorithm applied to Kartvelian Phonological Data: A Report on the LaDyCa Project (Language Dynamics in the Caucasus, IDEX Emergence, 2017–18): Methods in Theoretical and Empirical Dialectology”, paper presented at *Methods for Endangered Kartvelian Varieties Documentation & Analysis according to Language Dynamics Theory II*, 23 February 2018, Tbilisi: Ilia State University.
- Shanidze (1957): Акаки Шанидзе, “Принципы классификации грузинских диалектов [Principles of the Classification of Georgian Dialects]”, *Proceedings of the Joint Scientific Session of the Academy of Sciences of the Transcaucasian Republics on Social Sciences*, Baku, 828–832.

### Online resources

- Georgian Dialect Corpus (GDC): <http://corpora.co>  
Georgian National Corpus (GNC): <http://gnc.gov.ge>  
TITUS Database: <https://titus.uni-frankfurt.de>

# დიალექტოლოგიური ლექსიკოლოგიური მონაცემთა ბაზის აგება ქართული ენის ზიარი კოგნატების ანალიზისათვის

ელენე ჟერარდინი (პარიზი)

DOI: <https://doi.org/10.62235/dk.3.2024.8511>  
[helene.gerardin@inalco.fr](mailto:helene.gerardin@inalco.fr) || ORCID: [0009-0003-7406-3901](https://orcid.org/0009-0003-7406-3901)

წინამდებარე ნაშრომში განხილულია ის პრობლემები და გამოწვევები, რომელიც ჩვენ მიერ განხორციელებული პროექტის - ქართული დიალექტების ლექსიკურ მონაცემთა ბაზის შექმნას ახლდა თან. ეს გამოწვევები, უპირველეს ყოვლისა, დაკავშირებული იყო მონაცემთა შეგროვებისა და ფორმატირების მეთოდოლოგიური მიდგომების შემუშავებასთან, რომელიც დაწვრილებით არის განხილული სტატიაში.

პროექტის იდეა გაჩნდა IDEX-ის ერთ-ერთი პროექტის Linguistic Dynamics of the Caucasus (LaDyCa) ფარგლებში, რომელიც 2017-2018 წლებში განხორციელდა სორბონის უნივერსიტეტისა (პარიზი) და ილიას სახელმწიფო უნივერსიტეტის (თბილისი) თანამშრომლობით. თავად IDEX-ი („Excellence Initiatives“, ფრანგულად „Initiatives d' Excellence“) წარმოადგენს საფრანგეთის სახელმწიფოს მიერ შექმნილი პროგრამის „ინვესტიციები მომავლისთვის“ (“Investments for the Future”) ნაწილს, რომელიც მიზნად ისახავს კონკურენტუნარიანი მულტიდისციპლინური უმაღლესი საგანმანათლებლო და კვლევითი ინსტიტუტების შექმნას ქვეყანაში.

ქართული ენა წარმოდგენილია ფართო ენობრივი და გეოგრაფიული ვარიანტების სახით, რაც საფუძვლად უდევს ქართული დიალექტების მრავალფეროვნებას. სხვადასხვა ავტორის მიხედვით (იხ. მაგალითად Chikobava 1952, Shanidze 1957, Dzidziguri 1970 და Jorbenadze 1989), ქართულში გამოიყოფა 15–დან 20-მდე დიალექტი ან ქვედიალექტი, რომელიც კლასიფიკაციის განსხვავებულ პრინციპებს ეფუძნება.

პროექტი მიზნად ისახავდა ქართულ დიალექტებში დადასტურებული საერთო ლექსიკური მნიშვნელობის მქონე სიტყვათა შეგროვებასა და ლექსიკურ მონაცემთა ბაზის შექმნას ღია რესურსის სახით, რომელიც გამოსადეგი იქნებოდა არა მარტო დიალექტოლოგიური მასალის არქივირებისა და შემდგომი განვითარებისათვის, არამედ ქართულის ენობრივი ვარიაციების შესასწავლად, ასევე ქართული დიალექტების კლასიფიკაციის დასახვეწად და ენის დიაქრონიის შესწავლისა და რეკონსტრუქციის გასაადვილებლად. გარდა ამისა, ქართული დიალექტების ონლაინარქივირება შემდგომში მისი გაფართოვებისა და დახვეწის შესაძლებლობასაც იძლევა.

ქართული დიალექტების ლექსიკურ მონაცემთა ბაზის ბირთვს წარმოადგენს ბაზის სტრუქტურა, რომელიც აერთიანებს ქართული დიალექტების საერთო ლექსიკური მნიშვნელობის მქონე სიტყვებს (საზიარო კოგნატებს). ლექსიკური ერთეულების ბაზაში გაერთიანებისას მონაცემები, ერთი მხრივ, გაიფილტრა ნასესხები სიტყვებისაგან, რაც განსაკუთრებით უხვად დასტურდება სხვა



ენებთან კონტაქტში მყოფი დიალექტების შემთხვევაში; მეორე მხრივ, დადგინდა ლექსიკური ერთეულების წარმომადგენლობითი სია, არანაკლებ ასი ელემენტის მოცულობისა.

იმ სირთულეების გათვალისწინებით, რომელიც გაჩნდა ლექსიკური ერთეულების შერჩევის პროცესში, ასევე დიალექტოლოგიური რესურსების სიმცირიდან გამომდინარე, ჩვენ მიერ შექმნილ ბაზაში თავი მოუყარეთ რესურსებს როგორც ზეპირმეტყველების, ისე ბეჭდური წყაროებიდან. უნდა აღინიშნოს, რომ ბეჭდური წყაროები ძირითადად მოიცავს ლექსიკონებსა და სიტყვების სიებს, რომლებიც ჩვეულებრივ თან ერთვის ტექსტების კრებულებს. ასეთი სიების შემთხვევაში პრობლემას გვიქმნიდა ის გარემოება, რომ სიაში ძირითადად შესულია ისეთი ლექსემები, რომლებიც ყველაზე მეტად არიან დაშორებული სტანდარტს, მაშინ როდესაც საზიარო ლექსიკური ერთეულები ხშირად საერთოდ არ არის ასახული ლექსემათა ჩამონათვალში. გარდა ამისა, მონაცემებს ვაგსებდით გამოქვეყნებულ ტექსტებსა და ონლაინკორპუსებში დაცული მასალებით.

მონაცემების შეგროვების შემდეგ რესურსები გადავიდა მონაცემთა ბაზაში და განთავსდა ცხრილებში, სადაც ლექსიკური ერთეულები კლასიფიცირებულია დიალექტის მიხედვით. თითოეულ ერთეულს თან ახლავს თარგმანი ინგლისურ და ფრანგულ ენებზე. ლექსიკური ერთეულები წარმოდგენილია სტანდარტული ქართული ლემის მიხედვით და დალაგებულია ანბანური თანმიმდევრობით. შედარებისა და დიაქრონიული ანალიზის გასაადვილებლად, ლექსიკურ ერთეულებს ჩვენ ასევე დავურთეთ შესაბამისი ფორმები ძველ ქართულში და რეკონსტრუირებული ქართველური ეტიმონები (ასეთის არსებობის შემთხვევაში). გარდა ამისა, ლექსიკურ მონაცემთა ბაზის გასაფართოვებლად მასალას დამატა სხვა ქართველური ენების - მეგრულის, ლაზურისა და სვანურის ლექსიკური ერთეულები, რომლებიც ცხრილში განთავსდა ქართული დიალექტების მარჯვნივ განლაგებულ სვეტებში.

მონაცემთა ბაზის მიმდინარე ვერსიის სტრუქტურის აღწერის შემდეგ სტატიაში განხილულია მისი გამოყენების შესაძლებლობები და უპირატესობა ლინგვისტური კვლევებისათვის. სტატიაში წარმოდგენილია ლექსიკურ მონაცემთა ბაზაში ლევენშტაინის დისტანციების გამოთვლების შედეგად მიღებული კვლევის შედეგები და დიალექტების დიაქრონიული და დიატოპიური შედარება.

ლევენშტაინის ალგორითმის შედეგად მიღებული სტატისტიკური შედეგები, რომელიც პროექტის წევრის, ლაურა პიკარდის მიერ იქნა გამოანგარიშებული, ცალსახად გვიჩვენებს, რომ ქართული დიალექტები უფრო ახლოს არიან ერთმანეთთან, ვიდრე ეს ჩვეულებრივ შეინიშნება სხვა ენათა დიალექტების შედარებისას. კოეფიციენტებს შორის განსხვავება კი ემპირიულად ადასტურებს ქართველი დიალექტოლოგების დასკვნებს, რომ აღმოსავლურ და დასავლურ დიალექტებს შორის აშკარაა არსობრივი სხვაობა (იხ. დიაგრამა 5-ზე წარმოდგენილი იერარქიული კლასტერული დენდოგრამა).

აქ აუცილებლად უნდა აღინიშნოს ორი დიალექტის პერიფერიული პოზიციონირების ფაქტი; კერძოდ, ყველაზე გამოკვეთილი სხვაობა დასტურდება აზერბაიჯანში გავრცელებულ ინგილოურ დიალექტსა და აღმოსავლური მთის

დიალექტ თუშურში, მაშინ როდესაც ირანში გავრცელებული ფერეიდნული დიალექტი (ერთადერთი დიალექტი, რომელსაც სამი საუკუნის განმავლობაში არ ჰქონია შეხება სტანდარტულ ქართულთან), არ ავლენს ასეთ განსხვავებას. შეიძლება აქ გადამწყვეტი ფაქტორი იყოს თუშური დიალექტის კონტაქტი სხვა მთის დიალექტებთან, რაც მოითხოვს დამატებით კვლევას, თუ რა განასხვავებს მთის დიალექტებს ერთმანეთისაგან.

სტატიაში წარმოდგენილი დასკვნები ეყრდნობა ლევენშტაინის დისტანციების გაზომვის მეთოდს. თუმცა, უნდა აღვნიშნოთ, რომ ლევენშტაინის ალგორითმს გარკვეული ხარვეზები გააჩნია (იგი ფოკუსირებულია სიტყვის საწყის სეგმენტზე, მაშინ როდესაც ქართველური ენებისათვის გაცილებით მნიშვნელოვანია სიტყვის ბოლო ნაწილი). ამიტომ საინტერესო იქნებოდა მასალის ანალიზი დისტანციების სხვა ალგორითმის, მაგალითად იარო ვინქლერის დისტანციების ალგორითმის გამოყენებით. მითუმეტეს, რომ ჩვენ მიერ შექმნილი ქართული დიალექტების ლექსიკურ მონაცემთა ბაზა შეიძლება ადვილად იქნეს გამოყენებული სხვა ტიპის დისტანციების გამომთვლელი ალგორითმებისათვის.

სამომავლოდ ჩემი მიზანია შევავსო ქართული დიალექტების ლექსიკურ მონაცემთა ბაზა და შევქმნა ონლაინარქივი. რაც შეეხება გრძელვადიან მიზნებს, ქართული დიალექტების ლექსიკურ მონაცემთა ბაზის საფუძველზე უნდა შეიქმნას ღია წვდომის რესურსი, რომელიც შემდგომში თანდათან შეივსება და გაფართოვდება.

# The Linguistic Model of the Georgian Language (Synthesis of Nominal word-forms)

Ketevan Datukishvili, Nana Loladze, Merab Zakalashvili (Tbilisi)

**DOI:** <https://doi.org/10.62235/dk.3.2024.8512>

[ketevan.datukishvili@tsu.ge](mailto:ketevan.datukishvili@tsu.ge) || ORCID: [0009-0009-2104-556X](https://orcid.org/0009-0009-2104-556X)

[nana.loladze@tsu.ge](mailto:nana.loladze@tsu.ge) || ORCID: [0009-0005-1817-9963](https://orcid.org/0009-0005-1817-9963)

[merabza@gmail.com](mailto:merabza@gmail.com) || ORCID: [0009-0006-0757-6856](https://orcid.org/0009-0006-0757-6856)

**Abstract:** In this paper we present a unique linguistic model designed for the morphological synthesis and analysis of the Georgian language. The morphological structure of the Georgian language is fully covered by this model. Based on it, we have created a software with a special tool, namely, a processor enabling the generation and analysis of word forms. We would like to emphasise the fact that no particular grammatical theory is used in our linguistic model. Rather, the language data is provided in a different format and in a structured way, taking into account the theories currently in use. Our model as presented in this paper is a collection of morphological equations (often around 3,000 units) required to generate every word form from a single stem. Currently, more than 266 million word-forms can be synthesised utilising the Georgian language's morphological processor. While not all of these forms can be found in electronic texts or extensive corpora, they are all viable. These words have been referred to as “potential forms”. The possible forms are crucial for studying a variety of topics pertaining to natural language processing. Additionally, they aid in the resolution of specific artificial intelligence challenges. Beyond, we intend to pinpoint the exact frequencies of the data produced by the processor and specify the domains in which they are utilised.

**Keywords:** Linguistic model, Synthesis of word-forms, Morphological processor

With the aim of morphological synthesis and analysis of the Georgian language, we have created a special linguistic model. This model is a complete representation of the morphological structure of the Georgian language. Based on the given model, we have created a software with a special tool, namely, a processor enabling the generation and analysis of word forms.

In our linguistic model, the language is not described based on a certain grammatical theory. Instead, taking into account the existing theories, the language data are presented in a different format in a formalised way. The paper analyses the model and offers a corresponding mechanism of synthesis.

Each morpheme in this model has a symbolic marker. The unity of these markers yields a morphological formula of a word. Our method is based on the expression of words by means of symbolic formulae. This enables the representation of all types of word-forms in a formalised way.

The model reflects all the nominal word-forms derived from the same root, in particular, case-forms like მთას ('to the mountain' (dative)), მთებობს ('of the mountains' (genitive), etc.; forms with postpositions like მთაზე ('on the mountain'), მთისთვის ('for the mountain'); forms with particles like მთასაც ('also the mountain'), მთაშივე ('in the mountain itself'); forms with auxiliary verbs like მთაა ('is a mountain'), მთაშია ('is in the mountain'); forms with indirect speech particles like მთამ-მ ('he/she said that the mountain'), მთაში-მეთქი ('I said, in the mountain'); and other word-forms.

Based on the table of morphemes, each affix is expressed by means of a special symbol, for instance:  $\text{႗႗} - A_1$ ;  $\text{႔} - A_2$ ;  $\text{႕} - A_3$ ;  $\text{႖} - B_2$ ;  $\text{႗} - B_3$  etc. Any word-form can be expressed by means of these symbols. The unity of the symbols yields the morphological formula of a word. For example:  $\text{႗႗႗႗-႗႗-႗႗}$  ('children', ergative) –  $R + A_1 + B_3$ ;  $\text{႗႗႗႗-႗႗-႗-႗႗႗႗}$  ('for the child') –  $R + B_6 + C_1 + D_7$ . R is the root, whereas other symbols denote affixes (markers of number, case, postposition and so on).

The expression of words by means of morphological formulae is our method to enable the presentation of all kinds of words in a formalised way. The morphological formula is a pattern, based on which the program can make a synthesis of word-forms from the root.

Certain types of formulae are united in a model. A model is a unity of morphological formulae necessary for the generation of all word-forms from one stem (on the average, about 3,000 units). The phonological part of the model describes the changes related to the nominal inflections and their positions. The program also offers a list of phonological types.

Based on the above-mentioned linguistic model, we have created a special software tool: a morphological processor of the Georgian language. It consists of two components: a database and the program part. Based on the model and the phonological type, the synthesiser generates inflectional forms of the nominal parts of speech. As a result, the program yields a complete list of the nominal forms, including case-forms, nominal forms with postpositions, forms with particles and so on. This list can be arbitrarily referred to as a complete paradigm of the noun which, as we have mentioned, embraces approximately 3,000 units. Currently, the statistical data of the morphological processor are as follows:

- lemmas – 113,860
- formulae generated by the synthesiser – 916,893
- word-forms generated by the synthesiser – 266,478,980

Currently, using the morphological processor of the Georgian language, it is possible to implement the synthesis of over 266 million word-forms. Not all of these forms are found in electronic texts or ample corpuses, yet, they are potentially possible. We have termed these words as “potential forms”.

The “potential forms” are important for the research of diverse issues related to the processing of natural languages. In addition, they help solve certain tasks of artificial intelligence. We also plan to identify precise frequencies of the data generated by means of the processor and define the fields of their usage.

## ქართული ენის ლინგვისტური მოდელი (სახელური სიტყვაფორმების სინთეზი)

ქეთევან დათუკიშვილი, ნანა ლოლაძე, მერაბ ზაკალაშვილი

(ივანე ჯავახიშვილის სახელობის თბილისის სახელმწიფო უნივერსიტეტი)

DOI: <https://doi.org/10.62235/dk.3.2024.8512>

[ketevan.datukishvili@tsu.ge](mailto:ketevan.datukishvili@tsu.ge) || ORCID: [0009-0009-2104-556X](https://orcid.org/0009-0009-2104-556X)

[nana.loladze@tsu.ge](mailto:nana.loladze@tsu.ge) || ORCID: [0009-0005-1817-9963](https://orcid.org/0009-0005-1817-9963)

[merabza@gmail.com](mailto:merabza@gmail.com) || ORCID: [0009-0006-0757-6856](https://orcid.org/0009-0006-0757-6856)

### 1. შესავალი

ბუნებრივ ენათა დამუშავების ერთ-ერთ მნიშვნელოვან ამოცანას წარმოადგენს მორფოლოგიური ანალიზი და სინთეზი, რისთვისაც სხვადასხვა მეთოდი გამოიყენება. ენის დამუშავებისას უპირატესად ჯერ ხორციელდება მონაცემთა ანალიზი და მასზე დაყრდნობით – სინთეზი. ჩვენ თავდაპირველად შევასრულეთ სინთეზის ამოცანა და შემდეგ, მისი ალგორითმის საფუძველზე – ანალიზი. ჩვენი სინთეზის პროგრამა ეფუძნება ქართული ენის ლინგვისტურ მოდელს. ფლექსიის მოდელის ავტორია ქეთევან დათუკიშვილი, დერივაციის მოდელისა – ნანა ლოლაძე. პროცესორით სინთეზირებული ფორმების ვერიფიცირება ხორციელდება ტექსტების დიდ მასივებში (კორპუსებსა და სხვა ელექტრონულ წყაროებში).

სტატიაში განვიხილავთ ქართული ენის სახელის ლინგვისტურ მოდელს, რომლის მიხედვითაც სრულდება სახელურ სიტყვაფორმათა სინთეზი. მოდელი უზრუნველყოფს ერთი ფუძისგან ყველა იმ ფორმის გენერირებას, რომლებიც თეორიულად (ლინგვისტური წესების შესაბამისად) დასაშვებია თანამედროვე ქართულ ენაში.

ქართული, როგორც აგლუტინაციური ენა, აფიქსთა დიდი რაოდენობითა და ფორმაწარმოების მრავალფეროვნებით გამოირჩევა, ამიტომ განსაკუთრებით მნიშვნელოვანია იმგვარი ინსტრუმენტების შექმნა, რომლებიც უზრუნველყოფს ქართული ენის სინთეზსა და ანალიზს მორფოლოგიის დონეზე. ამ ამოცანის გადაწყვეტის არაერთი ცდა არსებობს: მოირერი (2007), მარგველანი (2008), ანთიძე (2009), ლოპუანიძე (2021)... ავტორები უპირატესად იყენებენ ენის მორფოლოგიური ანალიზის სტანდარტულ მეთოდებს, კერძოდ, სასრული პოზიციის ავტომატებს. ეს არის საყოველთაოდ ცნობილი მეთოდი ბუნებრივ ენათა მორფოლოგიური ანალიზისათვის,

მაგრამ გარკვეულ მოდიფიკაციას მოითხოვს იმ ენათა სისტემების აღწერისას, რომლებიც ფორმაწარმოების მრავალფეროვნებით ხასიათდება.

ჩვენ საკითხის გადაწყვეტის სხვა გზა ავირჩიეთ: ქართული ენისათვის შევქმენით ლინგვისტური მოდელი. მასში აღწერილია ენის ფლექსიური და დერივაციული სისტემები, აგრეთვე მორფოლოგიური და მორფოსინტაქსური პროცესები. ამ ინფორმაციის საფუძველზე განხორციელდა პროგრამული უზრუნველყოფა: შეიქმნა ინსტრუმენტი – მორფოლოგიური პროცესორი, რომლითაც შესაძლებელია ერთი ძირიდან სხვადასხვა ფუძის წარმოება დერივაციისა და კომპოზიციის საშუალებით. მაგალითად: ძირიდან „ძმა“ გენერირდება შემდეგი ფუძეები: დერივაციით – მოძმე, სამძო, უძმო, ძმობილ, ძმურ და მისთ. კომპოზიციით – და-ძმა, დედისძმა, ნახევარძმა, ცოლისძმა, ძმაბიჭ, ძმადნაფიც, ძმაკაც, ძმისშვილ და მისთ. თითოეული ფუძისაგან კი გენერირდება სახელური და ზმნური ფორმები. ზემოთ წარმოდგენილ ყველა ფუძეს აქვს სახელის პარადიგმა, ზოგი ფუძისაგან კი (ძმ, ძმობილ, ძმაბიჭ...) იწარმოება აგრეთვე ზმნური პარადიგმები. ამგვარად წარმოდგენილი მონაცემები საშუალებას იძლევა განხორციელდეს ქართული ენის სრული მორფოლოგიური სინთეზი – აღიწეროს როგორც ფლექსიური, ისე დერივაციული სისტემები.

სტატიაში წარმოდგენთ ლინგვისტური მოდელის შექმნის ჩვენეულ მეთოდს, კერძოდ, განვიხილავთ სახელის მოდელსა და მასზე დაფუძნებული სინთეზის მექანიზმს. აღვწერთ პროცესორის სტრუქტურას.

## 2. ქართული ენის სახელის ლინგვისტური მოდელი.

ჩვენს ლინგვისტურ მოდელში ენა არ არის აღწერილი რომელიმე კონკრეტული გრამატიკული თეორიის მიხედვით, არამედ არსებული თეორიების გათვალისწინებით ენობრივი მონაცემები წარმოდგინეთ განსხვავებული ფორმატით – ფორმალური წესების საშუალებით აღიწერა ქართული ენის ფლექსიური და დერივაციული შესაძლებლობანი, ფონოლოგიური პროცესები, მორფოსინტაქსური თავისებურებანი და სხვ.

მანქანური დამუშავებისათვის შექმნილ ლინგვისტურ მოდელში გარკვეულ შემთხვევებში, კომპიუტერული დამუშავების სპეციფიკის გამო, ტრადიციული გრამატიკებისაგან განსხვავებულად აღვწერთ ენობრივ მოვლენებს. მაგალითად, მო-



დელ ში სახელურ ფორმათა რიგში განიხილება შედგენილი შემასმენლები, რომლებიც წარმოადგენს ერთ სიტყვას, მაგრამ მოიცავენ ორ კომპონენტს: სახელურ ნაწილსა და დამხმარე ზმნის ხმოვან ელემენტს, მაგალითად:

ა) სახლია = სახლი + ა

სახლი-ი – სახ. ბრუნვის ფორმა

ა (<- არის) – მეშველი ზმნა

ბ) სახლისაა = სახლისა + ა

სახლი-ის-ა – ნათ. ბრუნვის ფორმა, გაგრცობილი

ა (<- არის) – მეშველი ზმნა

გ) სახლშია = სახლში + ა

სახლ-ში – თანდებულიანი სახელი

ა (<- არის) – მეშველი ზმნა და ა. შ.

ტრადიციულ გრამატიკებში, ბუნებრივია, ეს სიტყვები სახელებში არ განიხილება, მაგრამ ჩვენ მათ სახელურ ფორმათა გვერდით წარმოვადგენთ მანქანური დამუშავების სპეციფიკის გამო: ანალიზატორმა უნდა იცნოს ერთ სიტყვად გაფორმებული ნებისმიერი ერთეული, ამისათვის კომპიუტერულ მოდელში ასახული უნდა იყოს ყველა ამგვარი მონაცემი. ზემოთ დასახელებულ შედგენილ შემასმენლებს (სახლი-ა, სახლისა-ა, სახლში-ა...) ზმნურ პარადიგმებში ვერ განვიხილავთ, რადგან ზმნური ელემენტი აქ არის მხოლოდ -ა, სიტყვის მთავარი ნაწილი კი სახელური ფორმაა (ბრუნვის ფორმა, თანდებულიანი ან ნაწილაკიანი სახელი). ამიტომ ამგვარ ერთეულებს ჩვენ სახელური პარადიგმების წევრებად წარმოვადგენთ, მიუხედავად იმისა, რომ ისინი ერთგვარი შესიტყვებებია (უფრო სწორად, მომდინარეობს შესიტყვებისაგან) და არა სახელური სიტყვაფორმები.

ტრადიციული გრამატიკებისაგან განსხვავებული აღწერაა წარმოდგენილი ჩვენს მოდელში სხვა შემთხვევებშიც.

## 2.1. მოდელში ასახული სახელური ფორმები

მოდელში ასახულია ყველა სახელური ფორმა, რომლებიც ერთი ფუძიდან შეიძლება იწარმოოს, კერძოდ, ესენია:

**1. ბრუნვის ფორმები (გაუგრცობელი):**

<b>მხ. რ.</b>	<b>მრ. რ. (ებიანი)</b>	<b>მრ. რ. (ნართანიანი)</b>
სახ. მთა	მთ-ებ-ი	მთა-ნ-ი
მოთხ. მთა-მ	მთ-ებ-მა	მთა-თა
მიც. მთა-ს	მთ-ებ-ს	მთა-თა
ნათ. მთ-ის	მთ-ებ-ის	მთა-თა

და ა. შ.

**2. გაგრცობილი ბრუნვის ფორმები:**

<b>მხ. რ.</b>	<b>მრ. რ.</b>
მიც. მთას-ა	მთებს-ა
ნათ. მთის-ა	მთების-ა

და ა. შ.

**3. თანდებულიანი ფორმები:**

<b>მხ. რ.</b>	<b>მრ. რ.</b>
მთა-ზე	მთებ-ზე
მთის-თვის	მთების-თვის
მთ-ამდე	მთებ-ამდე

და ა. შ.

**4. გაგრცობილი თანდებულიანი ფორმები:**

<b>მხ. რ.</b>	<b>მრ. რ.</b>
მთისთვის-ა	მთებისთვის-ა
მთამდის-ა	მთებამდის-ა

და ა. შ.

**5. ნაწილაკდართული ფორმები:**

- ნაწილაკდართული ბრუნვის ფორმები:  
მთამა-ც, მთასა-ც, მთას-ღა... მთებსა-ც...
- ნაწილაკდართული თანდებულიანი ფორმები:  
მთაზე-ც, მთისთვისა-ც, მთისგან-ვე, მთებში-ღა...

**6. გაგრცობილი ნაწილაკდართული ფორმები:**

მთაც-ა, მთისაც-ა, მთაზეც-ა, მთამდეც-ა...

**7. მეშველზმნიანი ფორმები:**

- **მეშველზმნიანი ბრუნვის ფორმები:**  
მთა-ა, მთასა-ა, მთითა-ა...
- **მეშველზმნიანი თანდებულიანი ფორმები:**  
მთაში-ა, მთაზე-ა, მთასთანა-ა...
- **მეშველზმნიანი ნაწილაკიანი ფორმები:**  
მთაცა-ა, მთებილა-ა, მთაშიცა-ა...

**8. სხვათა სიტყვის ნაწილაკის დართვით მიღებული ფორმები:**

- **სხვათა სიტყვის ნაწილაკის დართვით მიღებული ბრუნვის ფორმები:**  
მთამ-ო, მთის-მეთქი, მთებით-თქო...
- **სხვათა სიტყვის ნაწილაკის დართვით მიღებული თანდებულიანი ფორმები:**  
მთაში-ო, მთაზე-მეთქი, მთასთან-თქო...
- **სხვათა სიტყვის ნაწილაკის დართვით მიღებული ნაწილაკიანი ფორმები:**  
მთაცა-ო, მთებილა-მეთქი, მთაშიც-თქო...
- **სხვათა სიტყვის ნაწილაკის დართვით მიღებული მიღებული მეშველ-ზმნიანი ფორმები:** მთაა-ო, მთებია-მეთქი, მთაშია-თქო...

როგორც წარმოდგენილი ნიმუშებიდან ჩანს, ქართულ ენაში სახელები მრავალფეროვანი ფორმაწარმოებით გამოირჩევა. ერთ სიტყვაფორმაში შეიძლება შეგვხვდეს ერთი, ორი, სამი და ა. შ. მორფემა. მორფემათა მაქსიმალური რაოდენობა არის 10. წარმოვადგენთ ნიმუშებს სხვადასხვა რაოდენობის მორფემათა შემცველი სახელებისთვის:

1. **ერთი მორფემა** – მთა
2. **ორი მორფემა** – მთა-ს  
ფუძე (მთა) + ბრუნვის ნიშანი (ს)
3. **სამი მორფემა** – მთ-ის-გან  
ფუძე (მთ) + ბრუნვის ნიშანი (ის) + თანდებული (გან)
4. **ოთხი მორფემა** – მთ-ის-ა-გან  
ფუძე (მთ) + ბრუნვის ნიშანი (ის) + სავრცობი (ა) + თანდებული (გან)
5. **ხუთი მორფემა** – მთ-ებ-ის-ა-გან

ფუძე (მთ) + რიცხვის ნიშანი (ებ) + ბრუნვის ნიშანი (ის) + სავრცობი (ა) + თანდებული (გან)

6. ექვსი მორფემა – მთ-ის-ა-გან-ა-ც

ფუძე (მთ) + ბრუნვის ნიშანი (ის) + სავრცობი (ა) + თანდებული (გან) + სავრცობი (ა) + ნაწილაკი (ც)

7. შვიდი მორფემა – მთ-ებ-ის-ა-გან-ა-ც

ფუძე (მთ) + რიცხვის ნიშანი (ებ) + ბრუნვის ნიშანი (ის) + სავრცობი (ა) + თანდებული (გან) + სავრცობი (ა) + ნაწილაკი (ც)

8. რვა მორფემა – მთ-ებ-ის-ა-გან-ა-ც-ო

ფუძე (მთ) + რიცხვის ნიშანი (ებ) + ბრუნვის ნიშანი (ის) + სავრცობი (ა) + თანდებული (გან) + სავრცობი (ა) + ნაწილაკი (ც) + სხვათა სიტყვის ნაწილაკი (ო)

9. ცხრა მორფემა – მთ-ებ-ის-ა-გან-ა-ც-ა-ა

ფუძე (მთ) + რიცხვის ნიშანი (ებ) + ბრუნვის ნიშანი (ის) + სავრცობი (ა) + თანდებული (გან) + სავრცობი (ა) + ნაწილაკი (ც) + სავრცობი (ა) + მეშველი ზმნა (ა)

10. ათი მორფემა – მთ-ებ-ის-ა-გან-ა-ც-ა-ა-ო

ფუძე (მთ) + რიცხვის ნიშანი (ებ) + ბრუნვის ნიშანი (ის) + სავრცობი (ა) + თანდებული (გან) + სავრცობი (ა) + ნაწილაკი (ც) + სავრცობი (ა) + მეშველი ზმნა (ა) + სხვათა სიტყვის ნაწილაკი (ო)

ჩვენს ლინგვისტურ მოდელში ასახულია სხვადასხვა მორფემის კომბინაციით მიღებული ყველა სახელი. როგორც ცნობილია, ფუძეებს აფიქსური მორფემები დაერთვის განსაზღვრული წესების მიხედვით. ამასთან, ზოგჯერ სახელურ ფორმებში ხდება ფონოლოგიური ცვლილებები. ეს კანონზომიერებანი აღწერილია შესაბამისი წესებისა და მონაცემების სახით, რომლებიც განთავსებულია მოდელის სხვადასხვა კომპონენტში.

## 2.2. სახელის მოდელის კომპონენტები.

სახელის მოდელი შედგება შემდეგი კომპონენტებისაგან:

- მორფემათა რანგობრივი ცხრილები
  - სიტყვაფორმათა შაბლონები
  - ფონოლოგიური ტიპების სია
- განვიხილავთ ამ კომპონენტებს.

### 2.2.1. მორფემათა რანგობრივი ცხრილი.

მართალია, ქართულ ენაში სახელის ფლექსიურ აფიქსთა რაოდენობა მნიშვნელოვნად დიდია, მაგრამ ისინი განაწილებულია მკაცრი რანგობრივი პრინციპით, კერძოდ:

- სიტყვაფორმებში მორფემათა მიმდევრობა წარმოდგენილია რანგების მიხედვით;
- სიტყვაფორმაში დასტურდება ამა თუ იმ რანგის მხოლოდ ერთი ელემენტი.

ნიმუშად წარმოვადგენთ სახელურ მორფემათა ცხრილის ფრაგმენტს.

N	R	A	B	C	D	E	F	G	H	I
1		-ებ	-0	-ა	-ში	-ა	-ც	-ა	-ა	-მეთქი
2		-ნ	-ო	-ო	-ზე		-ღა			-თქო
3		-თ	-მა		-დან		-ვე			-ო
4			-მ		-იდან					
5			-ს		-მდე					
6			-ის		-ამდე					
7			-ს		-თვის					

ცხრილი 1. სახელურ მორფემათა ცხრილის ფრაგმენტი

R-ფუძე, A-რიცხვის ნიშანი, B-ბრუნვის ნიშანი, C-სავრცობი1, D-თანდებული,

E-სავრცობი2, F-ნაწილაკი, G-სავრცობი3, H-მეშველი ზმნა, I-სიტყვა-სიტყვითი ნაწილაკი

ცხრილი 1-ის შესაბამისად, მოდელში თითოეული მორფემა გამოისახება სიმბოლური ნიშნით, მაგალითად:

ებ – A<sub>1</sub>

ნ – A<sub>2</sub>

თ – A<sub>3</sub>

ო – B<sub>2</sub>

მა – B<sub>3</sub>

მ – B<sub>4</sub>

ს – B<sub>5</sub> და ა. შ.

ცხრილში მორფემები წარმოდგენილია ენის მანქანური დამუშავების სპეციფიკის გათვალისწინებით და ზოგ შემთხვევაში აფიქსთა აღწერა არ ემთხვევა გრამატიკებში წარმოდგენილ მონაცემებს. განვიხილავთ ამგვარ შემთხვევებს.

როგორც ცნობილია, ქართული ენის სახელურ ფორმებში დასტურდება სავრცობი -ა. ტრადიციულ გრამატიკებში განიხილება ის აფიქსები, რომელთა გავრცობაც ხდება; კერძოდ, ცნობილია სავრცობის გამოყენების პოზიციები: ბრუნვის ნიშნის შემდეგ: სახლ-ს-ა, სახლ-ის-ა... თანდებულის წინ ან შემდეგ: სახლის-ათვის, სახლის-თვის-ა... ნაწილაკის წინ ან შემდეგ: სახლს-ა-ც, სახლში-ც-ა... და ა. შ.

მართალია, -ა ერთი მორფოლოგიური ოდენობაა (სავრცობი), მაგრამ მორფემათა რანგობრივ ცხრილში იგი წარმოდგენილია სამ სხვადასხვა სვეტში (C, E, G) იმის მიხედვით, თუ რომელი აფიქსია გავრცობილი, ე. ი. -ა რანგობრივად რომელ პოზიციას იკავებს სხვა მორფემებთან მიმართებით:

- თუ გავრცობილია ბრუნვის ნიშანი (B რანგის მორფემა), მაშინ  $a = C_1$ : ბავშვ-ს-ა;
- თუ გავრცობილია თანდებული (D რანგის მორფემა), მაშინ  $a = E_1$ : ბავშვ-თან-ა;
- თუ გავრცობილია ნაწილაკი (F რანგის მორფემა), მაშინ  $a = G_1$ : ბავშვი-ც-ა.

-ა სავრცობის ამგვარი დიფერენცირება მნიშვნელოვანია, რადგან ერთ ფორმაში შეიძლება მონაწილეობდეს ორი ან სამი -ა და თითოეულ მათგანს უნდა ჰქონდეს შესაბამისი სიმბოლური ნიშანი, რომ პროგრამამ შეძლოს ფორმის სინთეზი (ან ანალიზი). მაგალითად:

ბავშვ-ის-ა-თვის-ა-ც-ა – ამ ფორმაში პირველი  $a = C_1$ , მეორე  $a = E_1$ , და მესამე  $a = G_1$ .

უნდა აღინიშნოს ისიც, რომ ზოგჯერ სხვადასხვა სუფიქსი ფორმალურად ემთხვევა ერთმანეთს. ასეთ შემთხვევებში ისინი ორჯერ არის შეტანილი ცხრილში მათი ფუნქციური სხვაობის დასაფიქსირებლად. მაგალითად, -ს არის როგორც მიცემითი, ისე ნათესაობითი ბრუნვის ნიშანი, ცხრილში ისინი ამგვარად აისახება:

მიცემითი ბრუნვის ნიშანი:  $s = B_5$

ნათესაობითი ბრუნვის ნიშანი:  $s = B_7$

მაშასადამე, თითოეულ სახელურ მორფემას ცხრილში შეესაბამება თავისი სიმბოლური ნიშანი. ნებისმიერი სიტყვაფორმა შეიძლება გამოისახოს ამ ნიშნებით, რომელთა ერთობლიობა ქმნის სიტყვის მორფოლოგიურ ფორმულას. მაგალითად:



ბავშვ-ებ-მა –  $R + A_1 + B_3$

ბავშვ-ის-ა-თვის –  $R + B_6 + C_1 + D_7 \dots$

მორფოლოგიური ფორმულებით გამოსახვა არის ჩვენ მიერ შემოთავაზებული მეთოდი, რომლითაც შესაძლებელია ყველა ტიპის სიტყვაფორმის ფორმალური ბუნი სახით წარმოდგენა.

ფორმულით გამოისახება ნებისმიერი გრამატიკული ფორმა. მაგალითად:

ა) მოთხრობითი ბრუნვის ფორმის ფორმულა არის:  $R + B_3$ .

ბ) მრავლობით რიცხვში ნათესაობითი ბრუნვის ფორმის ფორმულა არის:  $R + A_1 + B_6$

გ) -ზე თანდებულებიანი ფორმის ფორმულა არის:  $R + D_2$  და ა. შ.

მორფოლოგიურ ფორმულებში  $R$  (ფუძე) არის ცვლადი ერთეული, ფორმულის ყველა დანარჩენი კომპონენტი არის უცვლელი (განსაზღვრული) მონაცემი. ერთი და იმავე ფორმულით სხვადასხვა ფუძისგან შეიძლება მივიღოთ კონკრეტული გრამატიკული ფორმები. მაგალითად, თუ ავიღებთ მოთხრობითი ბრუნვის ფორმულას ( $R + B_3$ ) და ვცვლით  $R$ -ს, მივიღებთ მოთხრობითი ბრუნვის ფორმებს.

მოთხრობითი ბრუნვის ფორმულა:  $R + B_3$

$B_3 =$  -მა

ფორმულაში  $R$ -ის ცვლით მიღებული სიტყვაფორმები:

$R =$  კაც: კაც-მა

$R =$  სახლ: სახლ-მა

$R =$  ბავშვ: ბავშვ-მა...

თუ სახელი ხმოვანფუძიანია, მაშინ მოთხრობითი ბრუნვის ფორმულა იქნება ამგვარი:  $R + B_4$

$B_4 =$  -მ

ამ ფორმულით სხვადასხვა ფუძისგან მიიღება შემდეგი სიტყვაფორმები:

$R =$  მთა: მთა-მ

$R =$  დედა: დედა-მ

$R =$  სპილო: სპილო-მ...

თუ ავიღებთ -ზე თანდებულებიანი სახელის ფორმულას ( $R + D_2$ ) და ვცვლით  $R$ -ს, მივიღებთ -ზე თანდებულებიან ფორმებს:

-ზე თანდებულებიანი სახელის ფორმულა:  $R + D_2$

$D_2 =$  -ზე

ფორმულაში R-ის ცვლით მიღებული სიტყვაფორმები:

R = კაც: კაც-ზე

R = სახლ: სახლ-ზე

R = ბავშვ: ბავშვ-ზე

R = მთა: მთა-ზე

R = დედა: დედა-ზე

R = სპილო: სპილო-ზე...

მაშასადამე, მორფოლოგიური ფორმულა არის გარკვეული ყალიბი, რომლის მიხედვით პროგრამას შეუძლია ამა თუ იმ ფუძიდან შესაბამისი სიტყვაფორმების სინთეზი.

### 2.2.2. სიტყვაფორმათა შაბლონები

შაბლონი წარმოადგენს იმ მორფოლოგიური ფორმულების ერთობლიობას, რომლებიც საჭიროა ერთი ფუძიდან ყველა სიტყვაფორმის დასაგენერირებლად, მაშასადამე, სრული პარადიგმის მისაღებად. შაბლონში სიტყვაფორმათა ნიმუშები ჩაწერილია ფორმულების სახით.

ცნობილია, რომ ქართულ ენაში სახელთა გარკვეული ჯგუფები ერთმანეთისაგან განსხვავდება ფორმაწარმოების თვალსაზრისით. შესაბამისად, გამოიყოფა შაბლონთა სხვადასხვა ტიპი. ისინი დამოკიდებულია სახელის ფუძეზე და პრაქტიკულად ემთხვევა ბრუნების ტიპებს, რომლებსაც ქართულ ენაში სხვადასხვა მეცნიერი განსხვავებულად გამოყოფს: შანიძე (1980), ჩიქობავა (1998), ონიანი (2003), უთურგაიძე (2009), არაბული (2016), ქირია (2022) და სხვ. ტიპებად დაყოფის პრინციპი განსხვავებულია: ფონოლოგიური, მორფოლოგიური, მორფონოლოგიური და ა. შ. მოდელისათვის რელევანტურია ფუძეზე დართულ მორფემათა აღწერა, რადგან მათი ერთობლიობა ქმნის ყალიბს (შაბლონს), რომლის საფუძველზეც უნდა განხორციელდეს სიტყვაფორმათა გენერაცია. ამიტომ ტიპიზაციისას ჩვენ ვითვალისწინებთ მორფოლოგიურ პრინციპს: შაბლონებს გამოვეყოფთ იმის მიხედვით, თუ რომელ აფიქსებს (ბრუნვის ნიშნებს, თანდებულებს, ნაწილაკებს...) დაერთავს სახელი.

ტრადიციულ გრამატიკებში კუმშვად-კვეცადი და ი-ბოლოხმოვნისანი უკვეცელი სახელები ცალკე ტიპებად არ განიხილება. ჩვენ მათ გამოვეყოფთ ცალკე შაბ-

ლონების სახით, რადგან ისინი განსხვავდება ყველა დანარჩენისგან ფორმაწარმოების (ფუძეზე დართული მორფემების) თვალსაზრისით. მართალია, სხვაობას ქმნის სულ რამდენიმე ფორმა, მაგრამ ამ უკანასკნელთა წარმოება შეუძლებელია სპეციალური შაბლონის გარეშე.

გვაქვს 5 ტიპის შაბლონი. წარმოვადგენთ მათ ცხრილის სახით. თითოეულ ტიპს ნიმუშად ვუწერთ ერთ-ერთ ამგვარ სახელს, რომელიც აზუსტებს, რა სახის სიტყვათა შაბლონია ის.

ბრუნვა	I ტიპი „კაცი“	II ტიპი „დედა“	III ტიპი „ფანჯარა“	IV ტიპი „წყარო“	V ტიპი „ტრამვაი“
სახელობითი	-ი	-Ø	-Ø	-Ø	-Ø
მთხრობითი	-მა	-მ	-მ	-მ	-მ
მიცემითი	-ს	-ს	-ს	-ს	-ს
ნათესაობითი	-ის	-ის	-ის	-ს	-ს
მოქმედებითი	-ით	-ით	-ით	-თი	-თ
ვითარებითი	-ად	-დ	-ად	-დ	-დ
წოდებითი	-ო	-Ø /-ვ/ -ო	-Ø /-ვ	-Ø /-ვ/ -ო	-Ø /-ვ/ -ო

ცხრილი 2. სახელის შაბლონთა ტიპები

წარმოდგენილი 5 შაბლონი აღწერს სახელთა ბრუნების ძირითად (რეგულარულ) ტიპებს; ამის გარდა არსებობს არარეგულარული ბრუნების ტიპები, რომლებსაც მიეკუთვნება ნაცვალსახელები; ამგვარი შემთხვევებისთვის ცალკე შაბლონები გამოიყოფა.

არარეგულარულ შაბლონებს ქმნიან აგრეთვე ე. წ. ფორმაუცვლელი სიტყვები (ზმნიხედები, თანდებულები, ნაწილაკები და შორისდებულები). როგორც ცნობილია, ეს მეტყველების ნაწილები ფორმაუცვლელ სიტყვათა ჯგუფს მიეკუთვნება, შესაბამისად, ტრადიციულ გრამატიკაში ისინი სახელურ ფლექსიაში არ განიხილება. აღნიშნული სიტყვები, მართალია, არ იბრუნვის, მაგრამ ქართულ ენაში ფორმაწარმოება გარკვეული სახით მათაც ახასიათებს, რადგან დაირთავენ ფლექსიურ აფიქსებს, კერძოდ, სავრცობს, თანდებულებსა და ნაწილაკებს. მაგალითად, ზმნიხედა აქ დაირთავს თანდებულებს: აქე-დან, აქა-მდე... ნაწილაკებს: აქა-ც, აქ-გე, აქ-და... ამიტომ ამ ტიპის სიტყვებიც მოდელში სახელური ფორმატით აღიწერება. მათთვის ცალკე შაბლონებია შექმნილი.

მოდელში სახელთათვის 20-ზე მეტი შაბლონია (რეგულარული თუ არარეგულარული) წარმოდგენილი. თითოეულ მათგანში ფორმულების სახით აღწერილია: ბრუნვის ფორმები, თანდებულებიანი ფორმები, ნაწილაკდართული ფორმები,

მეშველზმნიანი ფორმები და სხვათა სიტყვის ნაწილაკების დართვით მიღებული ფორმები.

ნიმუშად წარმოვადგენთ პირველი შაბლონის ფრაგმენტებს; მოდელში შაბლონი მოიცავს მხოლოდ ფორმულებს, აქ კი, ფორმულების უკეთ გააზრების მიზნით, ფრჩხილებში მითითებულია მათი შესაბამისი ნიმუშები (სიტყვაფორმები) ფუძისათვის **სახლ**.

### შაბლონი 1

#### ბრუნვის ფორმები

მხ. რ.	მრ. რ.		
სახ.	R+B <sub>2</sub> (სახლ-ი)	R+A <sub>1</sub> +B <sub>2</sub>	(სახლ-ებ-ი)
მოთხ.	R+B <sub>3</sub> (სახლ-მა)	R+A <sub>1</sub> +B <sub>3</sub>	(სახლ-ებ-მა)
მიც.	R+B <sub>5</sub> (სახლ-ს)	R+A <sub>1</sub> +B <sub>5</sub>	(სახლ-ებ-ს)
ნათ.	R+B <sub>6</sub> (სახლ-ის)	R+A <sub>1</sub> +B <sub>6</sub>	(სახლ-ებ-ის)
მოქ.	R+ B <sub>7</sub> (სახლ-ით)	R+A <sub>1</sub> +B <sub>7</sub>	(სახლ-ებ-ით)

და ა. შ.

#### თანდებულიანი ფორმები (გაუვრცობელი და გავრცობილი)

R+D <sub>1</sub>	(სახლ-ში)
R+D <sub>2</sub>	(სახლ-ზე)
R+D <sub>9</sub>	(სახლ-თან)
R+D <sub>9</sub> +E <sub>1</sub>	(სახლ-თან-ა)
R+B <sub>2</sub> +D <sub>10</sub>	(სახლ-ი-ვით)
R+B <sub>2</sub> +D <sub>10</sub> +E <sub>1</sub>	(სახლ-ი-ვით-ა)
R+D <sub>4</sub>	(სახლ-იდან)
R+D <sub>4</sub> +E <sub>1</sub>	(სახლ-იდან-ა)

და ა. შ.

#### ნაწილაკიანი ფორმები (გაუვრცობელი და გავრცობილი)

R+B <sub>2</sub> +F <sub>1</sub>	(სახლ-ი-ც)
R+B <sub>2</sub> +F <sub>1</sub> +G <sub>1</sub>	(სახლ-ი-ც-ა)
R+B <sub>3</sub> +F <sub>1</sub>	(სახლ-მა-ც)
R+B <sub>3</sub> +F <sub>1</sub> +G <sub>1</sub>	(სახლ-მა-ც-ა)
R+B <sub>5</sub> +C <sub>1</sub> +F <sub>1</sub>	(სახლ-ს-ა-ც)
R+B <sub>5</sub> +C <sub>1</sub> +F <sub>1</sub> +G <sub>1</sub>	(სახლ-ს-ა-ც-ა)
R+D <sub>1</sub> +F <sub>1</sub>	(სახლ-ში-ც)
R+D <sub>1</sub> +F <sub>1</sub> +G <sub>1</sub>	(სახლ-ში-ც-ა)
R+D <sub>2</sub> +F <sub>1</sub>	(სახლ-ზე-ც)

$R+D_2+F_1+G_1$	(სახლ-ზე-ც-ა)
$R+D_6+F_1$	(სახლ-ამდე-ც)
$R+D_6+F_1+G_1$	(სახლ-ამდე-ც-ა)

და ა. შ.

### მეშველზმნიანი ფორმები

$R+B_2+H_2$	(სახლ-ი-ა)
$R+B_5+C_1+H_2$	(სახლ-ს-ა-ა)
$R+B_7+C_1+H_2$	(სახლ-ით-ა-ა)
$R+A_1+B_2+H_2$	(სახლ-ებ-ი-ა)
$R+A_1+B_5+C_1+H_2$	(სახლ-ებ-ს-ა-ა)
$R+A_1+B_7+C_1+H_2$	(სახლ-ებ-ით-ა-ა)
$R+D_1+H_2$	(სახლ-ში-ა)
$R+D_2+H_2$	(სახლ-ზე-ა)
$R+D_4+E_1+H_2$	(სახლ-იდან-ა-ა)
$R+D_6+H_2$	(სახლ-ამდე-ა)
$R+D_9+E_1+H_2$	(სახლ-თან-ა-ა)

და ა. შ.

ამგვარი ფორმით (ფორმულების სახით) არის წარმოდგენილი შაბლონებში სახელური სიტყვაფორმები. რეგულარულ შაბლონებში ფორმათა რაოდენობა მერყეობს 736-იდან 756-ამდე. საშუალოდ – 746. ეს არის სახელები სხვათა სიტყვის ნაწილაკების (-მეთქი, -თქო, -ო) გარეშე. ყველა სიტყვას დაერთვის სამივე ნაწილაკი. საბოლოოდ, შაბლონში სიტყვაფორმათა სრული რაოდენობა არის 4-ჯერ მეტი, ე. ი. საშუალოდ 3 000 ერთეული.

მაშასადამე, ქართულ ენაში ერთი ფუძიდან დასაშვებია დაახლოებით 3 000 სახელური სიტყვის გენერაცია. რასაკვირველია, აქედან ყველა სიტყვაფორმა არ დასტურდება ხელმისაწვდომი ელექტრონული რესურსების ტექსტებში. ამის შესახებ ქვემოთ გვექნება მსჯელობა (იხ. 5. პოტენციური ფორმები).

### 2.2.3. ფონოლოგიური ტიპები.

მოდელის კიდევ ერთი კომპონენტია ფონოლოგიური ტიპების ჩამონათვალი. სახელთა ფორმაწარმოებისას მნიშვნელოვანია სახელურ ფუძეთა ფონოლოგიური მახასიათებელი, რომელიც განსაზღვრავს, იცვლება თუ არა ფუძე ფორმაწარმოების დროს და თუ იცვლება, რა ტიპის ცვლილებები ახასიათებს მას.

აქაც გვაქვს გარკვეული სხვაობა ტრადიციულ გრამატიკებში მოცემულ ფონოლოგიურ ტიპებსა და ლინგვისტურ მოდელში დაფიქსირებულ მონაცემებს შორის. მაგალითად, როგორც ცნობილია, არსებობს ხმოვანფუძიანი კვეცადი სახელე-ბი, რომელთა ბოლო ხმოვანი იკარგება (იკვეცება) მხოლოდითი რიცხვის ორ ბრუნ-ვასა (ნათესაობითი და მოქმედებითი) და ებ-იან მრავლობითში. მაგრამ ეს პროცესი ყველგან ერთგვაროვანი არ არის, კერძოდ: ხმოვანფუძიან სახელთა ნაწილი იკვე-ცება სამივე პოზიციაში (დედა: დედ-ის, დედ-ით, დედ-ებ-ი...), ნაწილი – მარტო მხო-ლოდით რიცხვში (სარკე: სარკ-ის, სარკ-ით, მაგრამ მრ. რ. – სარკე-ებ-ი...), ნაწილი კი – მხოლოდ ებიან მრავლობითში (გოგონა: გოგონა-სი, გოგონა-თი, მაგრამ მრ. რ. – გოგონ-ებ-ი...). ფონოლოგიური ცვლილების მხოლოდ ერთი ტიპი (კვეცადი) ვერ ასახავს ამ სხვაობას, ამიტომ ზემოთ დასახელებული ვარიანტებისთვის ჩვენ გამო-ვეყოფთ სამ სახეობას: კვეცადი მხოლოდითში, კვეცადი მრავლობითში და კვეცადი ორივე რიცხვში. ამავე მიზეზით ცალკე ტიპებად არის გამოყოფილი ორი სახეობა კუმშვად სახელებშიც: კუმშვადი ორივე რიცხვში (სოფელი: სოფლ-ის, სოფლ-ებ-ი...) და კუმშვადი მხოლოდითში (სასმელი: სასმლ-ის, მაგრამ მრ. რ. – სასმელ-ებ-ი...). ფონოლოგიური ტიპების ამგვარი დანაწევრება საშუალებას იძლევა, ზუსტად განხორციელდეს აღნიშნული ცვლილებები ფორმათა სინთეზის დროს.

წარმოვადგენთ ფონოლოგიურ ტიპებს:

- უცვლელი (კაცი, წყარო)
- კუმშვადი ორივე რიცხვში (სოფელი)
- კუმშვადი მხოლოდითში (სასმელი)
- ნაწილობრივ კუმშვადი (მინდორი)
- კუმშვად-თანხმოვანდამკარგველი (ამბავი)
- კვეცადი ორივე რიცხვში (დედა)
- კვეცადი მხოლოდითში (სარკე)
- კვეცადი მრავლობითში (გოგონა)
- კუმშვად-კვეცადი (ფანჯარა)

მოდელში აღწერილია, რა ცვლილებები ხორციელდება თითოეული ფონო-ლოგიური ტიპის შემთხვევაში და რომელ პოზიციებში. ამჯერად შემოვიფარგლე-ბით მხოლოდ ამ ტიპების დასახელებით, ფონოლოგიური ცვლილებების მოდელს სამომავლოდ ცალკე უფრო დეტალურად განვიხილავთ.



### **3. ქართული ენის მორფოლოგიური პროცესორი.**

ზემოთ განხილულ ლინგვისტურ მოდელზე დაფუძნებით შექმენით პროგრამული ინსტრუმენტი – ქართული ენის მორფოლოგიური პროცესორი. იგი შედგება ორი კომპონენტისგან: მონაცემთა ბაზისა და პროგრამული ნაწილისაგან.

**3.1. მონაცემთა ბაზა.** ბაზა მოიცავს ლინგვისტურ მოდელსა და ძირების ლექსიკონს.

**ლინგვისტურ მოდელში**, როგორც ზემოთ აღვნიშნეთ, ცხრილებისა და სიების სახით არის წარმოდგენილი სიტყვაფორმათა სინთეზისა და ანალიზისათვის საჭირო სხვადასხვაგვარი ენობრივი მონაცემები: მორფემათა რანგობრივი ცხრილები, ფონეტიკური ტიპების ჩამონათვალი, სახელისა და ზმნის პარადიგმათა შაბლონები, პირისა და რიცხვის ნიშნები, პირთა კომბინაციები და ა. შ.

მოდელის განთავსება მონაცემთა ბაზაში საშუალებას იძლევა, საჭიროების შემთხვევაში მოდელში განხორციელდეს ცვლილებები პროგრამული ჩარევის გარეშე. გარდა ამისა, ამგვარად ორგანიზებული მოდელის საფუძველზე სამომავლოდ ვფიქრობთ სპეციალური ინსტრუმენტის შექმნას, რომელიც გამოიყენება სხვადასხვა ენის მორფოლოგიური სინთეზისა და ანალიზის შესასრულებლად. იგი განსაკუთრებით მორგებული ინსტრუმენტი იქნება ავლუტინაციურ ენათა დამუშავებისას.

**ძირების ლექსიკონი** მოიცავს „ქართული ენის განმარტებითი ლექსიკონის“ რეატომეულს, აგრეთვე სხვადასხვა წყაროდან (ლექსიკონებიდან, კორპუსებიდან...) დამატებულ მასალას. ამ ეტაპზე ლექსიკონში დაფიქსირებულია 24 000-ამდე ძირი და მათგან ნაწარმოები 90 000-ზე მეტი ფუძე.

ლექსიკონში ძირები შეყვანილია შესაბამისი პარამეტრებით, კერძოდ, მიეთითება ძირის ფონეტიკური ტიპი და მისგან დერივაციული მოდელებით წარმოქმნილი ახალი ფუძეები, თითოეულ ფუძესთან დაფიქსირებულია მისგან მიღებულ სახელური და / ან ზმნური პარადიგმის შაბლონთა ტიპები, ფუძის ფონეტიკური ტიპი, ზმნური პარადიგმის შესაქმნელად საჭირო მონაცემები: ზმნის პირიანობა, გარდამავლობა და ა. შ.

სახელური ფორმების გენერაციისათვის ფუძეს მიეთითება ორი ინდექსი: შაბლონის ტიპი და ფონოლოგიური ტიპი. მაგალითად:

სახელის ფუძე	შაბლონის ტიპი	ფონოლოგიური ტიპი
მზე	„დედა“	კვეცადი მხოლოდობითში
ფურცელ	„კაცი“	კუმშვადი ორივე რიცხვში
სახლ	„კაცი“	უცვლელი
დაფა	„დედა“	კვეცადი ორივე რიცხვში
პეპელა	„ფანჯარა“	კუმშვად-კვეცადი
ენძელა	„წყარო“	კვეცადი მრავლობითში
კინო	„წყარო“	უცვლელი
ბგერა	„დედა“	კვეცადი ორივე რიცხვში

ცხრილი 3. ლექსიკონში განთავსებულ მონაცემთა ნიმუშები სახელის ფლექსიისათვის.

მონაცემთა ბაზის რედაქტორში ეს ინფორმაცია დაფიქსირებულია შემდეგნაირად:

### ფლექსიის რედაქტორი

ამ ფლექსიის წაშლა

სტატუსი დამონმებული

შემქმნელი keti

✔ მთლიანი დადასტურება

✘ მთლიანი უარყოფა

შედეგი სახლ-ი (სახლის, სახლები)

ლენა სახლი

ომონიმის სორტირების ნომერი

0

შენიშვნა

შენიშვნა

#### მთავარი პარამეტრები

კლასიფიკატორი

ქველი

ფლექსიის მოდელი

სახელის ფლექსია

სახელის პარადიგმა

კაცი

#### ფუძეები და არჩევითი მორფემები

1. ფუძე

სახლ (სახლ)

ფონეტიკური ტიპი

უცვლელი

სურათი 1. პროცესორში სახელის ფლექსიურ მონაცემთა შეყვანის ფორმატი

### 3.2. პროგრამული ნაწილი

პროგრამული ნაწილის კომპონენტებია: პროგრამა-რედაქტორი, სინთეზის პროგრამა (სინთეზატორი) და ანალიზის პროგრამა (ანალიზატორი).

პროგრამა-რედაქტორი არის მონაცემთა ბაზის მართვის სისტემა. ეს არის პროცესორის ის ნაწილი, რომელთანაც უშუალოდ უხდებათ მუშაობა ლინგვისტებს. იგი მოიცავს მოდელის რედაქტორსა და ძირების რედაქტორს. მათი საშუალებით შესაძლებელია ძირების ლექსიკონისა და მოდელის პარამეტრების შეყვანა, აგრეთვე მათი რედაქტირება: პარამეტრების შეცვლა ან ახლის დამატება.

მოდელსა და ძირების ლექსიკონში წარმოდგენილი მონაცემების საშუალებით სინთეზის პროგრამით (მორფოლოგიური სინთეზატორით) ხორციელდება სიტყვაფორმათა სინთეზი; სინთეზის ალგორითმის გამოყენებით ანალიზის პროგრამა (მორფოლოგიური ანალიზატორი) ახორციელებს უკუპროცესს: სიტყვაფორმათა ანალიზს.

### 4. სახელის სინთეზით მიღებულ სიტყვაფორმათა სიები (პარადიგმები).

შაბლონისა და ფონოლოგიური ტიპის მიხედვით სინთეზატორში ხდება სახელთა ფლექსიური ფორმების გენერაცია, რომლის შედეგად პროგრამა გვაძლევს სახელური ფორმების სრულ სიას; სია მოიცავს ბრუნების პარადიგმას, თანდებულის სახელებს, ნაწილაკიან სახელებს და ა. შ. პირობითად ამ სიას შეიძლება სახელის სრული პარადიგმა ვუწოდოთ. პარადიგმის ფორმებიდან ცალკე გამოიყოფა და სინთეზირებულ ფორმათა ჩამონათვალში ფიქსირდება ლემა (მხოლობითი რიცხვის სახელობითი ბრუნვის ფორმა), მონიშნება და ლემასთან ერთად მიეთითება აგრეთვე მხოლობითი რიცხვის ნათესაობითი ბრუნვისა და ებიანი მრავლობითის სახელობითი ბრუნვის ფორმები, რომლებიც მნიშვნელოვანია სახელის ფონოლოგიური ტიპის საილუსტრაციოდ.

ქართულ ლექსიკოგრაფიაში ამგვარი პრაქტიკა არსებობს: ქართული ენის განმარტებით ლექსიკონში სახელის ფონოლოგიური ტიპის საილუსტრაციოდ ფრჩხილებში მითითებულია ნათესაობითი ბრუნვის ფორმა, მაგრამ ხშირ შემთხვევაში (როდესაც მხოლობითი და მრავლობითი რიცხვის ფორმები განსხვავებულია კუმშვის ან კვეცის თვალსაზრისით) ეს ინფორმაცია საკმარისი არ არის, ამიტომ ჩვენ ვუთითებთ ორ მონაცემს: როგორც მხოლობითი რიცხვის ნათესაობითი ბრუნვის, ისე ებიანი მრავლობითის ფორმას. მაგალითად:

სარკე – სარკე (სარკის, სარკეები)

## სიტყვაფორმათა პარადიგმას ამგვარი სახე აქვს:

### პარადიგმა

გასუფთავება

შენახვა

- 1. სახლ
- 2. სახლი (ნიმუში 1)
- 3. სახლმა
- 4. სახლს
- 5. სახლსა
- 6. სახლის (ნიმუში 2)
- 7. სახლისა
- 8. სახლით
- 9. სახლითა
- 10. სახლად
- 11. სახლადა
- 12. სახლო
- 13. სახლები (ნიმუში 3)
- 14. სახლებმა
- 15. სახლებს
- 16. სახლებსა
- 17. სახლების
- 18. სახლებისა
- 19. სახლებით
- 20. სახლებითა
- 21. სახლებად
- 22. სახლებადა
- 23. სახლებო
- 24. სახლნი
- 25. სახლთ
- 26. სახლთა
- 27. სახლნო
- 28. სახლივით
- 29. სახლივითა
- 30. სახლსავით
- 31. სახლსავითა

Activate Windows

Go to Settings to activate Windows.

სურათი 2. სახელის სრული პარადიგმის ფრაგმენტი

პარადიგმაში შესაძლებელია სიტყვაფორმათა დაშლა აფიქსებად. ამისთვის არის სპეციალური დილაკი: „დეფისებით“:

### პარადიგმა

გასუფთავება

შენახვა

- 1. სახლ-
- 2. სახლ-ი- (ნიმუში 1)
- 3. სახლ-მა-
- 4. სახლ-ს-
- 5. სახლ-სა-
- 6. სახლ-ის- (ნიმუში 2)
- 7. სახლ-ისა-
- 8. სახლ-ით-
- 9. სახლ-ითა-
- 10. სახლ-ად-
- 11. სახლ-ადა-
- 12. სახლ-ო-
- 13. სახლ-ები- (ნიმუში 3)
- 14. სახლ-ებმა-
- 15. სახლ-ებს-
- 16. სახლ-ებსა-
- 17. სახლ-ების-
- 18. სახლ-ებისა-
- 19. სახლ-ებით-
- 20. სახლ-ებითა-
- 21. სახლ-ებად-
- 22. სახლ-ებადა-
- 23. სახლ-ებო-
- 24. სახლ-ნი-
- 25. სახლ-თ-
- 26. სახლ-თა-
- 27. სახლ-ნო-
- 28. სახლ-ი-ვით-
- 29. სახლ-ი-ვითა-
- 30. სახლ-სა-ვით-
- 31. სახლ-სა-ვითა-

Activate Windows

Go to Settings to activate Windows.

სურათი 3. მორფემებად დაშლილ სახელთა სრული პარადიგმის ფრაგმენტი

შესაძლებელია აგრეთვე სიტყვაფორმათა წარმოდგენა ფორმულების სახით:

**პარადიგმა**

გასუფთავება შენახვა

- 1. სახლ R1
- 2. სახლი R1-T1 (ნიმუში 1)
- 3. სახლმა R1-T3
- 4. სახლს R1-T4
- 5. სახლსა R1-T4-TA1
- 6. სახლის R1-T5 (ნიმუში 2)
- 7. სახლისა R1-T5-TA1
- 8. სახლით R1-T8
- 9. სახლითა R1-T8-TA1
- 10. სახლად R1-T10
- 11. სახლადა R1-T10-TA1
- 12. სახლო R1-T13
- 13. სახლები R1-S1-T1 (ნიმუში 3)
- 14. სახლებმა R1-S1-T3
- 15. სახლებს R1-S1-T4
- 16. სახლებსა R1-S1-T4-TA1
- 17. სახლების R1-S1-T5
- 18. სახლებისა R1-S1-T5-TA1
- 19. სახლებით R1-S1-T8
- 20. სახლებითა R1-S1-T8-TA1
- 21. სახლებად R1-S1-T10
- 22. სახლებადა R1-S1-T10-TA1
- 23. სახლებო R1-S1-T13
- 24. სახლნი R1-S2-T1
- 25. სახლოთ R1-S3
- 26. სახლოთა R1-S3-TA1
- 27. სახლონო R1-S2-T13
- 28. სახლივით R1-T1-U1
- 29. სახლივითა R1-T1-U1-TA1
- 30. სახლსავით R1-T4-TA1-U1
- 31. სახლსავითა R1-T4-TA1-U1-TA1

სურათი 4. სახელებისა და მათი შესაბამისი ფორმულების სრული პარადიგმის ფრაგმენტი არარეგულარული შაბლონები, ბუნებრივია, ფორმათა მცირე რაოდენობას შეიცავს. წარმოვადგენთ მათ ნიმუშსაც:

**პარადიგმა**

გასუფთავება შენახვა

- 1. გუმინ (ნიმუში 1)
- 2. გუმინა
- 3. გუმინაც
- 4. გუმინლა
- 5. გუმინვე
- 6. გუმინაა
- 7. გუმინლაა
- 8. გუმინვეა

სურათი 5. არარეგულარული პარადიგმის ფრაგმენტი

### 5. პოტენციური ფორმები

როგორც ზემოთ აღვნიშნეთ, ჩვენს მიზანს წარმოადგენდა ისეთი მოდელის შექმნა, რომელიც უზრუნველყოფს თანამედროვე ქართულ ენაში არსებული და პოტენციურად დასაშვები ნებისმიერი სიტყვაფორმის სინთეზს.

დღეისათვის მორფოლოგიური პროცესორის მონაცემთა სტატისტიკა ამგვარია:

- ლემები – 113 860
- სინთეზატორით დაგენერირებული ფორმულები – 916 893
- სინთეზატორით დაგენერირებული სიტყვაფორმები – 266 478 980

მაშასადამე, ბოლო მონაცემებით, ქართული ენის მორფოლოგიური პროცესორი აგენერირებს 266 მილიონზე მეტ სიტყვაფორმას. ყველა ეს ფორმა არ არის დადასტურებული ელექტრონულ ტექსტებში ან დიდი მოცულობის კორპუსებში, მაგრამ პოტენციურად დასაშვებია (ლინგვისტური მოდელი განსაზღვრავს მათ გენერაციას). ამგვარ სიტყვებს პირობითად „პოტენციურ ფორმებს“ ვუწოდებთ.

პოტენციური ფორმების დაფიქსირება პროცესორში ჩვენ საჭიროდ მიგვაჩნია. მოსალოდნელია, რომ დღეისათვის „პოტენციურად“ მიჩნეული ფორმები შეგვხვდეს რომელიმე ახალ (წერილობით ან ზეპირმეტყველების) ტექსტში. პროცესორმა უნდა შეძლოს მათი ამოცნობა. გარდა ამისა, საინტერესოა განისაზღვროს, მოდელით ნაგულისხმევი სიტყვაფორმებიდან რომელი ვარიანტებია რეალიზებული ენაში (ამასთან – რა სიხშირით) და რომელი – არა. ეს საკითხი ცალკე კვლევის საგანია. საინტერესოა აგრეთვე სხვადასხვა ენის მონაცემთა შედარება ამ თვალსაზრისით.

კვლევის შემდგომ ეტაპზე დიდი მოცულობის ტექსტური მონაცემებიდან ვგეგმავთ მასალის მოძიებას, რის საფუძველზეც განისაზღვრება სიტყვაფორმათა სიხშირული მახასიათებლები: აქტიურია ის, პასიური თუ პოტენციური. პოტენციური ფორმები ზოგიერთი ინსტრუმენტისათვის (მართლწერის შემმოწმებელი პროგრამები, სათარგმნი სისტემები და სხვ.) რელევანტური არ არის. მაგრამ ისინი მნიშვნელოვანია ბუნებრივ ენათა დამუშავების სხვადასხვა საკითხის კვლევისას, აგრეთვე ხელოვნური ინტელექტის ცალკეული ამოცანების გადასაწყვეტად. სამომავლოდ დაზუსტდება პროცესორით გენერირებულ მონაცემთა სიხშირული მახასიათებლები და განისაზღვრება მათი გამოყენების სფეროები.

## გამოყენებული ლიტერატურა

Meurer, P. (2007). A Computational Grammar for Georgian. Lecture Notes in Computer Science. Berlin: Springer.

ანთიძე, ჯ. (2009). ფორმალურ ენათა და გრამატიკათა თეორია, ბუნებრივი ენების კომპიუტერული მოდელირება, თბილისი: გამომცემლობა „ნეკერი“. ISBN 978-9941-404-92-4.

არაბული, ა. (2016). ახალი ქართული ენა, წიგნი I, სალიტერატურო ენის მორფოლოგია, გ. გოგოლაშვილისა და ა. არაბულის რედაქციით, თბილისი: „თბილისის სახელმწიფო უნივერსიტეტის გამომცემლობა“.

გოგოლაშვილი, გ. კვანტალიანი, ც. შენგელია, დ. (1989). ქართული ენის ზმნური ფუძეების ლექსიკონი, თბილისი: გამომცემლობა „მეცნიერება“.

გოგოლაშვილი, გ. კვანტალიანი, ც. შენგელია, დ. (1991). ქართული ენის სახელზმნური ფუძეების ლექსიკონი, თბილისი: გამომცემლობა „მეცნიერება“.

დათუკიშვილი ქ. (2020). ბრუნების ტიპებისათვის თანამედროვე ქართულში, იკე, XLVIII, თბილისი: „ივანე ჯავახიშვილის სახელობის თბილისის სახელმწიფო უნივერსიტეტის გამომცემლობა“ – <https://ice.tsu.ge/wp-content/uploads/2021/04/IKE-48-1-265-%E2%80%932020.pdf>

დათუკიშვილი, ქ. ლოლაძე, ნ. ზაკალაშვილი, მ. (2024). ქართული ენის ლინგვისტური მოდელი სიტყვაფორმათა სინთეზისა და ანალიზისთვის, მესამე საერთაშორისო კონფერენცია „საქართველო და კავკასია – წარსული, აწმყო, მომავალი“, თბილისი: „საქართველოს უნივერსიტეტის გამომცემლობა“.

დათუკიშვილი, ქ. ლოლაძე, ნ. ზაკალაშვილი, მ. (2024). დერივაციული მოდელები ქართული ენის მორფოლოგიურ პროცესორში, ენათმეცნიერ-კავკასიოლოგთა VII საერთაშორისო სიმპოზიუმის მასალები, თბილისი: „ივანე ჯავახიშვილის სახელობის თბილისის სახელმწიფო უნივერსიტეტის გამომცემლობა“. ISBN 978-9941-36-275-0

დათუკიშვილი, ქ. ლოლაძე, ნ. ზაკალაშვილი, მ. (2013). ქართული ენის მორფოლოგიური პროცესორი – მანქანური ანალიზისა და სინთეზის ინსტრუმენტი, ჟურნალი „ქართველოლოგია“, N4, თბილისი: „ილიას სახელმწიფო უნივერსიტეტის გამომცემლობა“.

ღობჯანიძე, ი. (2021). „ქართული ენის მორფოსინტაქსური ანოტირებისა და სასრული პოზიციის მორფოლოგიური ანალიზის პრინციპები“, თბილისი: „ილიას სახელმწიფო უნივერსიტეტის გამომცემლობა“. ISBN 978-9941-18-366-9

მარგველანი, ლ. (2008). ქართული ენის კომპიუტერული მოდელები, თბილისი: გამომცემლობა „ინტელექტი“. ISBN 5-7859-0078-5

ონიანი, ალ. (2003). თანამედროვე ქართული სალიტერატურო ენა, თბილისი: „სულხან-საბა ორბელიანის სახელობის თბილისის სახელმწიფო პედაგოგიური უნივერსიტეტის გამომცემლობა“.



- უთურგაიძე, თ. (2009). ქართული ენის დონეთა ძირითადი მახასიათებლების ურთიერთზე-  
მოქმედებისათვის გლობალურ ენობრივ სისტემაში, თბილისი: გამომცემლობა „მე-  
რიდიანი“. ISBN 978-9941-10-194-6
- ქართული ენის განმარტებითი ლექსიკონი. (1950 – 1964). ტ. I – VIII, თბილისი: „საქართვე-  
ლოს სსრ მეცნიერებათა აკადემიის გამომცემლობა“. –  
<https://ice.tsu.ge/liv/ganmartebiti.php>
- ქართული ენის განმარტებითი ლექსიკონი. (2008 – 2019). ტ. I – IV, თბილისი: გამომცემლობა  
„მერიდიანი“ – <https://ice.tsu.ge/liv/ganmartebiti.php>
- ქართული ენის ეროვნული კორპუსი – <http://gnc.gov.ge/gnc>
- ქართული ლექსიკონი – <https://www.ganmarteba.ge/>
- ქირია, ჯ. (2022). არსებითი სახელის ბრუნების პარადიგმული მოდელები თანამედროვე ქარ-  
თულში, „საენათმეცნიერო ძიებანი“, XIII, თბილისი. ISSN 1987-6653
- შანიძე, ა. (1980). ქართული ენის გრამატიკის საფუძვლები, თხზულებანი 12 ტომად, ტ. III,  
თბილისი: „თბილისის უნივერსიტეტის გამომცემლობა“.
- ჩიქობავა, არნ. (1998). ქართული ენის ზოგადი დახასიათება, თბილისი: გამომცემლობა  
„ქართული ენა“.
- ჯორბენაძე, ბ. ლოლაძე, ნ. კიკონიშვილი, მ. (2014). ქართული ენის სახელური ფუძეების  
ლექსიკონი, თბილისი: „თბილისის უნივერსიტეტის გამომცემლობა“. ISBN 978-  
9941-13-326-8

# Research into Equivalence in Multilingual Parallel Corpora

Manana Tandaschwili (Frankfurt), Mariam Kamarauli (Hamburg)

**DOI:** <https://doi.org/10.62235/dk.3.2024.8514>

[tandaschwili@em.uni-frankfurt.de](mailto:tandaschwili@em.uni-frankfurt.de) || ORCID: [0009-0005-7812-9124](https://orcid.org/0009-0005-7812-9124)

[mariam.kamarauli@uni-hamburg.de](mailto:mariam.kamarauli@uni-hamburg.de) || ORCID: [0009-0006-0404-4424](https://orcid.org/0009-0006-0404-4424)

**Abstract:** Shota Rustaveli’s epic “The Knight in the Panther’s Skin” is an exceptional monument of aphoristic style. The complexity of Shota Rustaveli’s worldview, which is mirrored in his epic, is a result of the mixing of mediaeval and Renaissance thinking. It combines Christian (and, in general, religious) thoughts with antique philosophy (Plato, Dionysius the Areopagite, Aristotle) and mythical and transcendental visions with analytical thinking. Rustaveli’s worldview, as well as several religious-philosophical doctrines – Christianity, Areopagitica, Neoplatonism, and Aristotelianism – are all reflected in the aphorisms used in the epic, which plays a special role not only in the cultural memory of the Georgians but also in the history of world literature. The epic has been translated into 58 languages and is included in the registry of the world cultural heritage of UNESCO. The variety of languages into which the epic has been translated provides a unique opportunity for the creation of the multilingual parallel corpus ‘Rustaveli goes digital’, which is an outstanding resource for translation studies to examine the quality of translations, analyse translation methods, pinpoint translators’ strategies, and contrast them. But how can we judge the quality of translations, especially in the case of aphorisms? Determining the degree of equivalence between the source and target texts is not sufficient for a qualitative evaluation of the translations of the epic; e.g., the translation quality of the aphoristic style in the translations is one of the challenges that we believe needs to be addressed. In the present paper, we discuss our research model on how to evaluate the translation quality of aphorisms.

**Keywords:** Digital Rustvelology, Shota Rustaveli, Translation Studies, Equivalence, Aphorism, Multilingual Parallel Corpus

## Introduction

Shota Rustaveli’s epic “The Knight in the Panther’s Skin” is a masterpiece of Georgian literature. It plays a special role not only in the cultural memory of the Georgian people but also in the history of world literature. The epic has been translated into 58 languages<sup>1</sup> and is included in the registry of the world cultural heritage of UNESCO.<sup>2</sup>

The importance of “The Knight in the Panther’s Skin” is indicated by the fact that nowadays, there are more than 150 translations of the poem in different languages. Some languages offer several translations of the epic by different translators (for example, in German, English, Spanish, French, Russian, Hebrew, Ossetian, Armenian, Persian, Azerbaijani, etc.). The time gap between the translations of the poem in some languages is more than 100 years, which greatly increases the area of interdisciplinary research of the epic. Considering this factor, the epic and its translations are an excellent resource for creating a multilingual parallel corpus. The concept for the creation of a multilingual parallel corpus in 58 languages was developed in 2018 at the Institute for Empirical Linguistics, Goethe University of Frankfurt.<sup>3</sup> The ‘Rustaveli goes digital’ corpus (RDG) now comprises 38 translations in 25 languages<sup>4</sup> and represents a

---

<sup>1</sup> Gippert 2024: 85–88.

<sup>2</sup> <https://en.unesco.org/silkroad/silk-road-themes/documentary-heritage/manuscript-collection-shota-rustavelis-poem-knight-panthers>.

<sup>3</sup> Initially by Manana Tandashvili, since 2020 the project is led by Mariam Kamarauli.

<sup>4</sup> <https://rustaveli-goes-digital.de/>.

unique opportunity to investigate topics in linguistics, literary studies, lexicology, philosophy, translation studies, history, sociology, ethnology, etc. in an interdisciplinary and cross-linguistic manner.

The epic itself is an exceptional monument of aphoristic style. The complexity of Shota Rustaveli's worldview is a result of the mixing of medieval and Renaissance thinking. It combines Christian (and, in general, religious) thought with antique philosophy (Plato, Dionysius the Areopagite, Aristotle), and mythical and transcendental visions with analytical thinking.<sup>5</sup> Rustaveli's worldview, as well as several religious-philosophical doctrines – Christianity, Areopagitica, Neoplatonism, Aristotelianism – are all reflected in the aphorisms.

First and foremost, a multilingual corpus like RGD is a resource for translation studies to examine the quality of translations, analyse translation methods, pinpoint translators' strategies, and contrast them. But how can one judge the quality of translations, especially in the case of aphorisms? Determining the degree of equivalence between the source and target texts is not sufficient for a qualitative evaluation of the epic's translations, e.g. the translation quality of the aphoristic style in the translations is one of the challenges that we believe needs to be addressed. In the present paper, we discuss our research model on how to evaluate the translation quality of aphorisms.

### **1. Theoretical framework of the research: How to evaluate equivalence in aphorisms on the base of multilingual corpora?**

Equivalency is a key factor in translation studies that is used to evaluate the quality of translations. The following factors should be the main considerations for evaluating equivalency: a) from a lexicographical point of view, equivalence is to be placed in an intermediate position between equivalence at the system level and at the text level;<sup>6</sup> b) equivalence must be at least partly corpus-based; c) equivalence cannot take into account all possible translation solutions found in the texts, but only the recurrent prototypical text equivalences; and d) equivalence must take into account the syntactic and semantic restrictions and preferences in both languages as well as the syntagmatics.

Lexicographic equivalence must be achieved by analysing the proto-typical behaviour of phraseological units at the discourse level, for which it is essential to be able to distinguish between typical uses and peripheral or non-representative uses. Furthermore, lexicographic equivalence differs from textual equivalence in that it does not attempt to consider all possible cases of translation of a phraseology from L1 to L2. In addition, it would be advisable for dictionaries to explain possible usage restrictions, combinatorial problems or divergences in the meaning structure of the respective phraseologisms in L2 in addition to the equivalences in L2.<sup>7</sup>

---

<sup>5</sup> Khintibidze 2018: 19.

<sup>6</sup> Cf. Mellado Blanco 2010: 177-178.

<sup>7</sup> Mellado Blanco (2015: 155): “La equivalencia lexicografica debe obtenerse mediante el analisis del comportamiento prototipico de las unidades fraseologicas en el nivel del discurso, para lo cual es esencial saber discriminar los usos tipicos de los perifericos o poco representativos. Por otra parte, la equivalencia lexicografica se diferencia de la equivalencia textual en que no pretende hacerse eco de toda la casuistica posible de traducciones de un fraseologismo de la L1 en la L2. Ademas de ello, en los diccionarios seria aconsejable que junto a las

Equivalence research in multilingual parallel corpora refers to the investigation of how meaning-related or semantically similar expressions in different languages can be compared with each other in order to evaluate the translation quality. The question of equivalence plays a particularly important role in the case of aphorisms, which are often succinct, concise and stylistically sophisticated statements, as this is not only a question of content, but also of the form, tone and effect of the expression. Various aspects must be considered, such as:

- a) **Semantic equivalence:** It is about the correspondence of meaning between an aphorism and its translation. As aphorisms are often formulated metaphorically or ironically, it can be difficult to grasp the exact meaning in another language.
- b) **Stylistic equivalence:** Aphorisms are often strongly stylised (e.g. through wordplay, rhyme or rhythm). Maintaining stylistic equivalence can be a challenge, as the linguistic means vary from language to language.
- c) **Cultural equivalence:** Many aphorisms are deeply rooted in one culture and may not be easily understood by speakers of another language. Cultural connotation plays a role here, going beyond the literal meaning.
- d) **Pragmatic equivalence:** This examines whether the intended effect of the aphorism (e.g. humorous, thought-provoking, critical) is retained in the target language.
- e) **Functional equivalence:** Establishing the equivalence of aphorisms from a functional point of view, e.g. Aphorisms can be didactic, instructive, explanatory, motivating, reprimanding, encouraging so on.

## 2. Methodology

The aim of this work is to investigate the equivalence of aphorisms using a multilingual parallel corpus. Such a corpus offers several advantages for equivalence research:

- a) **Possibilities for comparison:** One can directly compare the original language of the aphorism with its translations in different languages in order to analyse similarities or differences in meaning, style and effect.
- b) **Statistical analyses:** With the help of modern corpus linguistics methods, patterns in the translations can be identified, for example which linguistic structures are used particularly frequently to translate certain aphorisms.
- c) **Identification of translation strategies:** It can be analysed whether translators tend to translate aphorisms literally or whether they tend to make creative adaptations in order to achieve the desired effect in the target language.

Equivalence research of aphorisms in a multilingual parallel corpus can analyse the different levels of translation and show how aphorisms can have been understood and translated in different cultural and linguistic contexts. While researching the equivalence of aphorisms, various methodological approaches are used to analyse translation equivalences. Since aphorisms are particularly concise, stylistically demanding and often strongly context-bound,

---

equivalencias en la L2 se explicaran las posibles restricciones de uso, cuestiones combinatorias o divergencias en la estructura del significado de los respectivos fraseologismos en la L2.”

their analysis requires specialised methods that take both linguistic and cultural aspects into account. Below are some of the most important methodologies.

### 2.1. Contrastive analysis

Contrastive analysis systematically compares the structural and semantic differences and similarities between an aphorism in the source language and its translation. This method primarily analyses:

- **Lexical differences:** Which words are used in the different languages, and what nuances are lost or added?
- **Syntactic structures:** Do sentence structure and word order differ between the versions?
- **Stylistic elements:** How are linguistic devices such as alliteration, puns, metaphors or rhymes preserved or altered in the translations?

### 2.2. Functional equivalence analysis

The focus here is on how the intended function or effect of an aphorism is preserved in the target language. This method analyses whether the aphorism has the same effect on the reader in translation, for example with respect to:

- **Pragmatic effects:** Does the aphorism have the same ironic, humorous or thought-provoking effect in the target language?
- **Target group adequacy:** Is the aphorism formulated in a way that is understandable and relevant for the respective audience?

### 2.3. Cultural equivalence

Aphorisms are often deeply rooted in one culture and contain allusions that cannot easily be translated into another language. The analysis of cultural equivalence examines how translators deal with these cultural differences with respect to:

- **Cultural adaptation:** Is the aphorism directly translated or culturally adapted to achieve a similar effect?
- **Use of equivalents:** Do translators potentially find similar proverbs or idioms in the target language to achieve the same effect?

### 2.4. Translation strategy analysis

This method analyses the different strategies used by translators to translate aphorisms. Various approaches are analysed:

- **Literal translation:** The attempt to render the aphorism as accurately and faithfully as possible, even if stylistic or cultural aspects are lost.
- **Free translation:** The meaning of the aphorism is adapted to make it work in the target language, even if the exact wording is changed.
- **Paraphrasing:** An explanation or rephrasing of the aphorism to preserve the underlying meaning.

## 2.5. Corpus-based analysis

In equivalence research on aphorisms, a multilingual parallel corpus can be used to systematically compare aphorisms and their translations. Corpus linguistic tools and statistical methods help to identify translation patterns:

- **Collocation analysis:** Which typical words occur in connection with certain aphorisms, and how do they change in the translations?
- **Frequency analysis:** How often are certain structures or translation strategies used in a corpus?
- **Pattern recognition:** Recognition of recurring translation strategies or deviations between different target languages.

## 2.6. Descriptive translation studies

This method analyses which translations were actually used in practice and why. The aim is to describe the translation process and the decisions that translators take without making a normative judgement:

- **Researching translator decisions:** Why did translators choose a particular translation strategy?
- **Empirical studies:** How do translations of aphorisms differ over time or between different target languages?

## 2.7. Qualitative text analysis

In this method, aphorisms and their translations are analysed in a qualitative in-depth analysis. This involves a detailed analysis of the text, taking into account not only the language but also the historical and cultural context. This is particularly relevant when analysing aphorisms with strong cultural references or linguistic nuances.

## 2.8. Reception research

Another method is to analyse how aphorisms and their translations are understood and received by the respective target groups. This involves analysing

- **Reader reactions:** How do readers interpret the aphorism in the source and target languages?
- **Effect on the audience:** Does the aphorism have the same emotional or intellectual effect in the translation as in the original?

## 2.9. Error analysis

This method focuses on possible “errors” or misunderstandings in the translation. It analyses whether certain aspects of an aphorism have been lost or misunderstood in the translation. Equivalence research on aphorisms requires a combination of linguistic, cultural and stylistic analyses. By using these methodologies, translators and researchers can better understand how aphorisms function in different languages and what challenges arise when translating these concise and often complex linguistic expressions.

### 3. The question of equivalence in the case of aphorisms by Rustaveli

Before researching equivalence in case of aphorisms by Rustaveli, we must first solve three main problems: a) determine the number of aphorisms in the epic, b) develop criteria for determining the equivalence of aphorisms, and c) classify them in terms of structure, content and function.

#### 3.1. Number of aphorisms

Although numerous works have been dedicated to the aphorisms contained in “The Knight in the Panther’s Skin”, opinions of scholars still vary regarding their origin, their content, their structure and even their number in the epic. Currently, there are about 30 monolingual or bilingual editions of Rustaveli’s aphorisms, in which the number of examples varies from 24 to 234.<sup>8</sup> The main reason for this diversity is that there are no strict criteria for distinguishing between aphorisms and similar expressions (didactic phrases, parables, proverbs etc.).

In our research, we take a lower number (52) of aphorisms, because on the one hand, these aphorisms have been taken from the so-called academic edition of the epic (Rustaveli 1957) and, on the other hand, we agree with the editors that the identified aphorisms are by no means controversial.

#### 3.2. Structure of aphorisms

There are different definitions of aphorisms, but in terms of structure it is always emphasised that an aphorism should reflect wisdom in a short form. The uniqueness of Shota Rustaveli’s epic lies in its inclusion of aphorisms that not only follow a classical, concise structure (wisdom expressed in a single sentence) but also feature more intricate structures. We will provide a few examples here.

More than half of the 52 aphorisms (28 aphorisms) in our database exhibit a classic, simple structure: they are short. However, the other half of the aphorisms have a complex structure.

#### a) Aphorisms with a simple structure:<sup>9</sup>

No	Stanza	Aphorism
1	0.30	ავსა კაცსა ავი სიტყვა ურჩევნია სულსა, გულსა. <i>An evil man prefers an evil word to his soul, to his heart.</i>
2	1.39	ღვეკვი ღომისა სწორია, ძე იყოს თუნდა ზვადია. <i>The lion’s whelps are equal, be they male or female.</i>
3	1.50	რასაცა გასცემ, შენია; რას არა, დაკარგულია! <i>What you give away is yours; what you keep is lost.</i>
4	4.162	მარგალიტი არვის მიჰხედეს უსასყიდლოდ, უვაჭრელად. <i>A pearl does not fall into anyone’s lap without buying, without trading.</i>
5	4.164	მარტოება ვერას მიზამს, მცავს თუ ცისა ძაღთა დასი; <i>Loneliness cannot harm me if the heavenly powers (planets) protect me.</i>
6	17.432	კარგი საქმე კაცსა ზედა აზომ თურე არ წახდების. <i>A good deed that someone has done is apparently never lost.</i>
7	24.549	უხვად გავსცემდეთ, ვავსებდეთ, სიძუნწე უმეცრულია. <i>Let us spend generously, let us fill them (with treasures); avarice is foolishness.</i>

<sup>8</sup> Gabunia 2006: 141–142.

<sup>9</sup> Translations of aphorisms into English belong to the authors of this paper.



8	29.726	თუ ღხინი გვინდა ღმრთისაგან, ჭირიცა შევიწყნაროთა. <i>If we desire happiness from God, we must accept griefs also.</i>
9	30.736	რამცა სადა გაუმარჯვლა კაცსა, ფიცთა გამტეხელსა! <i>When and where did ever a breaker of oaths prevail?</i>
10	30.751	ზოგი თქმა სჯობს არა-თქმასა, ზოგი თქმითაც დაშავდების. <i>Sometimes speech is better than silence, sometimes by speaking we spoil (things).</i>
11	30.761	მტერი მტერსა ვერას ავნებს, რომე კაცი თავსა ივნებს. <i>A foe cannot hurt his foe as a man harms himself.</i>
12	30.773	ოდეს კაცსა დაეჭიროს, მაშინ უნდა ძმა და თვისი. <i>When a man is in trouble then needs he brother and kinsman.</i>
13	32.795	არვის ძალ-უც ხორციელსა განგებისა გარდავლენა. <i>Nobody in the flesh has the power the destiny to blot.</i>
14	32.799	სჯობს სახელისა მოხვეჭა ყოველსა მოსახვეჭელსა! <i>It is better to get glory than all goods.</i>
15	34.822	თუ თავი შენი შენ გახლავს, ღარიბად არ იხსენები. <i>If you have yourself, no one will consider you poor.</i>
16	36.875	ჭირსა შიგან გამაგრება ასრე უნდა, ვით ქვითკირსა. <i>In grief one should strengthen himself like a stone wall.</i>
17	36.877	არ იცი, ვარდნი უეკლოდ არავის მოუკრებიან! <i>Don't you know that no one ever plucked a rose without a thorn?</i>
18	36.883	ჰკითხე ასთა, ქმენ გულისა, რა გინდა ვინ რა გივაზიროს! <i>Enquire of a hundred, do what pleases your heart, in spite of what anyone may advise you.</i>
19	36.901	გველსა სვრელით ამოიყვანს ენა ტკბილად მოუბარი. <i>The sweetly discoursing tongue lures forth the serpent from its lair.</i>
20	36.904	გონიერთა მწვრთელი უყვარს, უგუნურთა გულსა ჰგმირდეს. <i>The intelligent love the teacher; the dumb take it as a stabbing in the heart.</i>
21	38.924	დიდი ღხინია ჭირთა თქმა, თუ კაცსა მოუხდებოდეს. <i>Great comfort it is to speak of troubles when a man has the opportunity.</i>
22	39.956	წელულსა დანა ვერა ჰკურნებს, გაჰკვეთს, ანუ გაამსივნებს <i>A knife cannot cure a wound; it cuts or causes a swelling.</i>
23	41.1046	დიდთა ხეთა მოერევის, მცირე დასწვავს ნაბერწკალი, <i>A tiny spark overcomes and burns up great trees.</i>
24	41.1050	ხამს ყოველისა დაჯერება, ბრძენი სჯერა მოწვეწადსა. <i>It is necessary to believe in all; a wise man has faith in the future.</i>
25	43.1083	მას მკურნალმან ვერა ჰკურნოს, თუ არ უთხრას, რაცა სჭირდეს. <i>The physician can't cure him who does not tell him what hurts him.</i>
26	47.1254	თუ ყვავი ვარდსა იშოვნის, თავი ბულბული ჰგონია. <i>If a crow finds a rose it considers itself a nightingale.</i>
27	52.1348	იგი მიენდოს სოფელსა, ვინცა თავისა მტერია! <i>He only can trust this world who is his own foe.</i>
28	52.1361	ბოროტსა სძლია კეთილმან, არსება მისი გრძელია! <i>Evil was defeated by Good, his essence is lasting.</i>

The quality of the translation of these aphorisms can be evaluated according to various aspects such as:

- a) the semantic equivalence of concepts in the given sentence
- b) the stylistic equivalence
- c) the syntactic and pragmatic equivalence (word order, informational structure)
- d) the functional equivalence (didactic, instructive, explanatory, motivating, re-primanding, encouraging and so on).

**b) Aphorisms with a complex structure:**

No	Stanza	Aphorism
29	1.35	რა ვარდმან მისი ყვავილი გაახშოს, დაამჟნაროსა, იგი წახდების, სხვა მოვა ტურფასა საბადნაროსა; When the rose's flower is dried and withered it falls, and another blooms in the lovely garden.
30	29.703	სამს მოყვარე მოყვრისათვის თავი ჭირსა არ დამრიდად, გული მისცეს გულისათვის, სიყვარული გზად და ხიდად; A friend should spare himself/herself no trouble for his/her friend's sake, He/she should give heart for heart, love as a road and a bridge.
31	32.798	არ-დავიწყება მოყვრისა აროდეს გვიზამს ზიანსა; გჰმობ კაცსა აუგიანსა, ცრუსა და ღალატიანსა! The mindfulness of a friend never harms us. I despise the man who is shameless, false, and treacherous.
32	34.815	მართლად უთქვამს მეცნიერთა: "წყენაო ჭირთა ბადე"; ესეგვარსა ნუოდეს იქმ, საქმე ხოლე გაიცადე. Truly say the sages: Spite is a net of woes! Never act in such a way, consider the matter carefully.
33	35.854	ესე არაკი მართალი ჩინს ქვასა ზედა სწერია: „ვინ მოყვარესა არ ეძებს, იგი თავისა მტერია“. This true saying is written on a stone in China: “Who does not seek a friend is his/her own foe.”
34	36.902	არას გარგებს სწავლეულება, თუ არა იქმ ბრძენთა თქმულსა: არ იხმარებ, რას ხელსა ჰხდი საუნჯესა დაფარულსა? Learning avails you not, if thou do not what the wise have said; What advantage to you is a hidden treasure if you will not use it?
35	44.1094	ბინდის გვარია სოფელი, ესე თურ ამად ბინდების, კოკასა შიგან რაცა დგას, იგივე წამოდინდების! The world is a kind of twilight, so here all is dusky. Whatever is in the pitcher, the same flows forth.
36	46.1191	სიკვდილამდის ვის მოუკლავს თავი კაცსა მეცნიერსა? რა მისჭირდეს, მაშინ უნდან გონებანი გონიერსა!" Until death what sensible man slays himself, When he needs it, his wit's the last thing that goes to a wise.
37	46.1211	მოყვარე მტერი ყოველისა მტრისაგანც უფრო მტერია; არ მიენდობის გულითა, თუ კაცი მეცნიერია. A friend-foe is the most hateful foe than any foe. He will not heartily confide if a man is wise.

Rustaveli uses aphorisms to provide sage advice in a clear, succinct way. In epic poetry, aphorisms with a complex structure are typically presented in a single stanza. The first three lines may have explanatory content and be expressed by theses, maxims or sentences, while the fourth line is a conclusion and can often be regarded separately as an aphorism with a simple structure. Nevertheless, 15 aphorisms of Rustaveli's have the form of whole stanza.

32.800 ვერ დაიჭირავს სიკვდილსა გზა ვიწრო, ვერცა კლდოვანი;  
მისგან გასწორდეს ყოველი, სუსტი და ძალ-გულოვანი;  
ბოლოდ შეყარეს მიწამან ერთგან მოყმე და მსცოვანი.  
სჯობს სიცოცხლესა ნაზრახსა სიკვდილი სახელოვანი!  
“Neither a narrow nor a stony path can hold back death.  
Through it all are levelled, weak and strong-hearted.

In the end the earth unites youth and greybeard in one place.  
Better a glorious death than a shameful life!”

Sometimes wisdom is conveyed allegorically in a stanza, the fourth line of which is an aphorism.

36.878 ვარდსა ჰკითხეს: “უგზომ ტურფა რამან შეგქმნა ტანად, პირად?

მიკვირს, რად ხარ ეკლიანი? პოვნა შენი რად არსი ჭირად?”

მან თქვა: "ტკბილსა მწარე ჰპოვებს, სჯობს, იქმნების რაცა ძვირად:  
ოღეს ტურფა გაიფუდეს, არღარა ღირს არცა ჩირად".

“The rose was asked: ‘Who made you so beautiful in the body, in the face?

I wonder why you are thorny, why it is so difficult/painful to find you!

She said: ‘You get to the sweet through the bitter; better is that which is hard to come by:

If the beauty is cheap, it is not even worth dried fruit. (= If turf becomes cheaper, it will lose its value).”<sup>10</sup>

In other cases, wisdom is conveyed in the form of maxims and sentences. In such instances, the whole stanza should be considered as an aphorism:

29.718 გული კრულია კაცისა, ხარბი და გაუძღომელი,

გული უამ-უამად ყოველთა ჭირთა მთმო, ღხინთა მნდომელი,

გული – ბრმა, ურჩი ხედვისა, თვით ვერას ვერ გამზომელი,

ვერცა ჰპატრონობს სიკვდილი, ვერცა პატრონი რომელი!

“The heart of man is accursed, greedy and insatiable.

Sometimes the heart desiring joys endures all griefs.

The heart is blind, disobedient in seeing, not at all able to measure itself.

Nor death itself can master it, also nor an owner.”

Since the aphorisms in Rustaveli’s epic show different structures (some are simple, others are complex), it is necessary to investigate the equivalence of aphorisms with different structures using different methods.

#### 4. Levels of equivalence research of aphorisms

The research on the aphorisms of the epic can be carried out on different levels. To do this, it is necessary to break down the aphorisms into their components. One unique aspect of the epic’s aphorisms is their varied structures, which vary from one-line aphorisms to intricate two-, three-, or four-line constructions as we have seen above. In order to create comparable standards for aphorism equivalence research, it is necessary to analyse and classify the aphorisms according to their structure. Accordingly, the aphorisms should be divided into the following units in order to do this:

- lexical components that are part of aphorism
- phrasal components that are part of sentence
- Logical components that are part of aphorisms.

<sup>10</sup> The equivalence of this aphorism, as well as the analysis of the idiom არღარა ღირს არცა ჩირად in English translations, has been discussed in detail by Mariam Gobianidze (2024: 91–97).

#### 4.1. Aphorisms with a simple structure

In this section, we will analyse one of the aphorisms listed above with a simple structure: ავსა კაცსა ავი სიტყვა ურჩევნია სულსა, გულსა “An evil man prefers an evil word to his soul, to his heart”. We shall analyse this sentence at several linguistic levels below.

##### a) Lexical analysis at the word level

There are four substantives in this sentence: კაცი ‘man’, სიტყვა ‘word’, სული ‘soul’, გული ‘heart’, along with one adjective, ავი ‘evil’, and one verb, ურჩევნია ‘to prefer’.

- კაცი has two main meanings in Georgian: the abstract meaning (‘human, man’) and the gender-specific meaning of human, namely ‘man’ (‘male human’). In the analysed sentence, this word is used with the first meaning and refers to ‘human’.
- The term სიტყვა ‘word’ has four meanings: 1) a lexical unit made up of phonemes; 2) speech; 3) report; and 4) expressed information. In the given context, it means ‘speech’.
- სული ‘soul’ has several meanings depending on its field of use: 1) a set of mental abilities of a person, e.g. ‘thinking’, ‘consciousness’; 2) the immaterial beginning, which is the driving force of the objects and events of nature, like ‘mind’ (in philosophy); 3) the inner world of a person, characteristic mental qualities of persons (ability, talent, character, feeling, feeling, inclinations...), which determine their actions, behaviours, points of view; 4) a disembodied supernatural being; 5) the divine side of the human being that opposes the body (flesh) and which is supposedly immortal. Because the epic depicts ancient philosophy on the one hand and religious thought on the other, here the word means the ‘immaterial beginning’, which is the driving force of the objects and events of nature, the ‘soul’, which is immortal in contrast to the body.
- გული ‘heart’: 1) the main vital organ of a human or animal that provides blood circulation throughout the body; 2) a symbol of human feelings, emotions and mood. When სული ‘soul’ and გული ‘heart’ occur simultaneously, the expression means the divine side of the human being that opposes the body (flesh) and which is supposedly immortal; the whole inner world of man, feeling and mind.
- ავი ‘evil, bad’ is an adjective with negative meaning and an antonym of ‘good’. When using this adjective with a noun, the entire NP assumes a negative meaning.
- ურჩევნია ‘to prefer, consider better’: this verb is semantically close to the verb ‘choose’. The morphosyntactic feature of this verb is that it requires two objects: a direct object and an object of comparison. The verb ‘prefer’ expresses the subject’s attitude toward, and preference for, a particular item.

##### b) Morphological analysis

The sentence consists of 7 words:

ავსა	კაცსა	ავი	სიტყვა	ურჩევნია	სულსა	გულსა
A.DAT.SG	N.DAT.SG	A.NOM.SG	N.NOM.SG	V.PRES.S3SG	N.DAT.SG	N.DAT.SG
<i>evil</i>	<i>human</i>	<i>Evil</i>	<i>word</i>	<i>to prefer</i>	<i>soul</i>	<i>heart</i>

The predication in the sentence is given by the verb ურჩევნია ‘to prefer’ which belongs to the semantic group of *verba sentiendi* and accordingly agrees with the subject in the dative case, which is a syntactic object but a logical subject (here a LS in form of the NP ავსა კაცსა ‘evil human’) and with the syntactic subject in the nominative case which is a logical object (here a LO in form of the NP ავი სიტყვა ‘evil speech’). The predication additionally requires an object of comparison on the morphosyntactic level. In this sentence, the object of comparison is presented by two nouns in the dative case: სულსა ‘soul’ and გულსა ‘heart’.

c) Syntactic analysis

Both the logical subject and the logical object are represented by nominal phrases containing the structure ADJECTIVE + NOUN:

ავსა კაცსა	ავი სიტყვა	ურჩევნია	სულსა	გულსა.
<i>evil human</i>	<i>evil word</i>	<i>prefers</i>	<i>soul</i>	<i>heart</i>
LS	LDO	V <sub>sentiendi</sub>	O <sub>COMP</sub>	O <sub>COMP</sub>

d) Information structure

Before looking at the informational structure, we must break down the aphorism into phrases. The information structure of the sentence shows the word order SOV:

<b>(L)S</b>	<b>(L)DO</b>	<b>V</b>	<b>O (of comparison)</b>
ავსა კაცსა	ავი სიტყვა	ურჩევნია	სულსა, გულსა
	<b>Foc</b>		

In this sentence, the LS is placed in the sentence-initial position. The subject is followed by the object and because it precedes the verb, it is focused. The transposition test shows that the focus is shifted when the sentence constituents are repositioned:

<b>(L)S</b>	<b>V</b>	<b>(L)DO</b>	<b>O (of comparison)</b>
ავსა კაცსა	ურჩევნია	ავი სიტყვა	სულსა, გულსა
<b>Foc</b>			

e) Stylistic analysis

It should be noted that the adjective ავი ‘evil’ in this aphorism specifies two nouns: კაცი ‘man’ and სიტყვა ‘word’. The use of the mentioned adjective in both cases is an artistic device and the author uses it specifically to give the phrase a melodic quality on the one hand and to add more intensity to the expression of the aphorism on the other hand.

The statement given in the aphorism has a very dynamic structure: the repetition of the adjective ავი ‘evil’ on the one hand, and the listing of the objects of comparison სული, გული (‘soul, heart’) on the other, gives rhythm to the aphorism and reinforces its emotional impact.

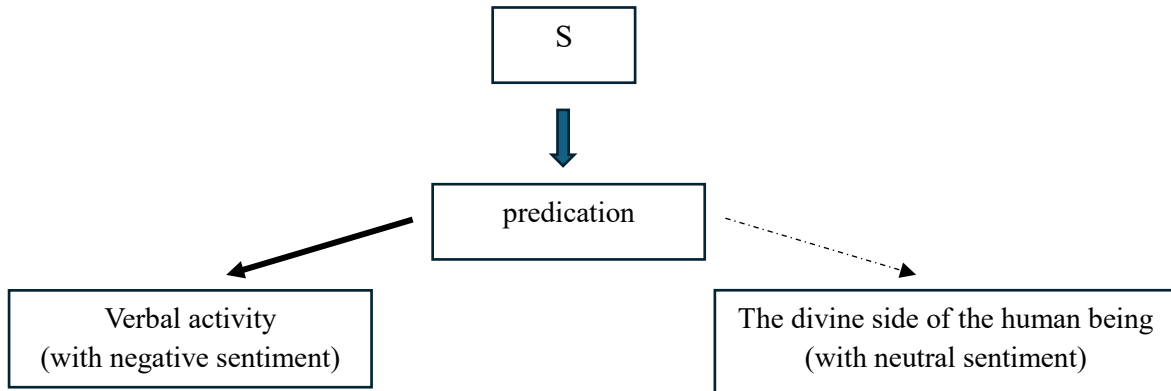
f) Sentiment analysis

The nouns კაცი and სიტყვა are neutral in terms of sentiment. Both are accompanied by the adjective ავი (ავი კაცი vs. ავი სიტყვა), which has a negative connotation. By using an adjective with negative connotation as an attribute, both NPs acquire negative sentiment as shown in Table I.

Word level	აგ-სა	კაც-სა	აგ-ი	სიტყვა	ურჩევნი-ა	სუფლ-სა	გულ-სა
Translation	evil	human	evil	word	to prefer	Soul	heart
Sentiment	NEG	NEU	NEG	NEU	NEU	NEU	NEU
Phrase level	აგსა კაცსა		აგი სიტყვა		ურჩევნია	სუფლსა	გულსა
Translation	evil human		evil word		Prefers	Soul	Heart
Sentiment	NEG		NEG		NEU	NEU	NEU

**Table I: Analysis of the sentence constituents on the different levels**

The predicate ურჩევნიად ‘to prefer’ morphosyntactically requires two objects – a direct object and an object of comparison as shown in Fig. 1:



**Fig. 1: Predication with two objects**

This contrasts two facets of the human being: the verbal activity (functionality) and the divine side of the human being. By choosing (favouring) between these two objects explicitly given lexically in the sentence, the negative and positive aspects are contrasted as shown in Table II.

Phrase level	აგსა კაცსა	აგი სიტყვა	ურჩევნია	სუფლსა	გულსა
Translation	evil human	evil word	prefers	soul	Heart
Sentiment	NEG	NEG	NEU	NEU	NEU
Sentence level	NEG				

**Table II: Sentiment analysis of the sentence**

The result of the sentiment analysis on the sentence level is negative: **an evil human prefers a bad thing to a good thing**. The semantics of the verb ‘to prefer’ on the one hand, and favouring a negative object over the neutral one on the other hand, gives the sentence a negative sentiment overall.

#### 4.2. The translations of the aphorism in other languages

Now let us take a look at how this aphorism has been translated into other languages. We will analyse the equivalents of the given aphorism in three German, two English, two French, one Russian and three Spanish translations.

HH *Hohn und Schmähung sind dem Unhold mehr als Leib und Leben lieb.*  
Mockery and vituperation are dearer to the fiend than life and limb.

MP *Böse Wörter über alles ein Verwegener nur liebt.*  
A temerarious man only loves evil words above all.

HB *Böses Wort liebt der Böse mehr als die eigene Seele.*

The evil one loves the evil word more than his own soul.

MW *But an evil man loves an evil word more than his soul or heart.*

LC *But an evil man loves an evil word more than he loves his soul.*

GB *L'homme méchant prise surtout une parole vénéneuse.*

The wicked man takes especially venomous speech.

STs *Une âme basse, au lieu du coeur, choisit la parole malsaine.*

A poor soul, instead of the heart, chooses unhealthy speech.

ShN *Злое слово больше жизни почитает лишь злодей.*

Only a villain honours an evil word more than life.

GT *Pero es cierto que un mal hombre ama a una palabra perversa más que a su alma y a su corazón.*

But it is true that an evil man loves a wicked word more than his heart and soul.

LM *Un alma baja escoge la palabra malsana en lugar del corazón.*

A low soul chooses the unwholesome word rather than the heart.

MB *El malvado ama las palabras venenosas más que el alma o el corazón.*

A wicked man loves poisonous words more than the soul or the heart.

There are four concepts in this aphorism: *man, word, soul, heart*. Table III illustrates with which equivalents these concepts are represented in the translations.

		კაცი	სიტყვა	სული	გული
GER	HH	<i>Unhold</i> 'fiend'	<i>Hohn und Schmähung</i> 'mockery and vituperation'	<i>Leben</i> 'life'	<i>Leib</i> 'body'
	MP	<i>Verwegener</i> 'temerarious man'	<i>Wörter</i> 'words'	<i>über alles</i> 'above all'	
	HB	<i>Böse</i> 'evil one'	<i>Wort</i> 'word'	<i>Seele</i> 'soul'	-
ENG	MW	<i>man</i>	<i>word</i>	<i>soul</i>	<i>heart</i>
	LC	<i>man</i>	<i>word</i>	<i>soul</i>	-
FR	GB	<i>homme</i> 'man'	<i>parole</i> 'word'	-	-
	STs	<i>âme</i> 'soul'	<i>parole</i> 'word'	-	<i>coeur</i> 'heart'
RU	ShN	<i>злодей</i> 'villain'	<i>слово</i> 'word'	<i>больше жизни</i> 'above life'	
ES	GT	<i>hombre</i> 'man'	<i>palabra</i> 'word'	<i>alma</i> 'soul'	<i>corazón</i> 'heart'
	LM	<i>alma</i> 'soul'	<i>palabra</i> 'word'	-	<i>corazón</i> 'heart'
	MB	<i>malvado</i> 'evil one'	<i>palabra</i> 'word'	<i>alma</i> 'soul'	<i>corazón</i> 'heart'

**Table III: Lexical equivalents of the concepts კაცი, სიტყვა, სული, გული**

Now we analyse the selected aphorism according to Nida's "Formal and Dynamic" model of translation to highlight the difficulties arising during the process of translation. Table IV shows the formal (F) and dynamic (D) equivalents in different colours.



		კაცი	სიტყვა	სული	გული
GER	HH	D	D	D	D
	MP	D	F	D	
	HB	F	F	F	-
ENG	MW	F	F	F	F
	LC	F	F	F	-
FR	GB	F	F	-	-
	STs	D	F	-	F
RU	ShN	F	F	D	
ES	GT	F	F	F	F
	LM	D	F	-	F
	MB	F	F	F	F

**Table IV: Formal and Dynamic equivalence of the concepts კაცი, სიტყვა, სული, გული**

The variety of the equivalents of word-level translations for კაცი ‘man’ is due to the fact that in the source text, the word კაცი ‘man’ is accompanied by the adjective ავი ‘evil’, which modifies the neutral content of the word into a negative one: ავი კაცი ‘evil man’. In the target language, translators often choose a noun with a negative content as an equivalent instead of a noun phrase consisting of adjective+noun: *Unhold* ‘fiend’ (HH), *Böse* ‘evil one’ (HB), *злодей* ‘villain’ (ShN), *malvado* ‘evil one’ (MB).

These examples show that equivalence research at the word level cannot be productive, and therefore we will divide the sentence into the phrases and continue researching equivalence at the NP level. Table V shows the Equivalents of the phrases ავი კაცი and ავი სიტყვა:

	NP 1	NP 2
	ავი კაცი	ავი სიტყვა
HH	<i>Unhold</i> ‘fiend’	<i>Hohn und Schmähung</i> ‘mockery and vituperation’
MP	<i>ein Verwegener</i> ‘a temerarious man’	<i>böse Wörter</i> ‘evil words’
HB	<i>der Böse</i> ‘the evil one’	<i>böses Wort</i> ‘evil word’
MW	<i>an evil man</i>	<i>an evil word</i>
LC	<i>an evil man</i>	<i>an evil word</i>
GB	<i>l’homme méchant</i> ‘the wicked man’	<i>une parole vénéneuse</i> ‘a poisonous word’
STs	<i>une âme basse</i> ‘a poor soul’	<i>la parole malsaine</i> ‘the unhealthy word’
ShN	<i>лишь злодей</i> ‘only/just a villain’	<i>злое слово</i> ‘evil word’
GT	<i>un mal hombre</i> ‘a bad man’	<i>a una palabra perversa</i> ‘a wicked word’
LM	<i>Un alma baja</i> ‘a lowly/poor soul’	<i>la palabra malsana</i> ‘the unhealthy word’
MB	<i>el malvado</i> ‘an evil man’	<i>las palabras venenosas</i> ‘poisonous words’

**Table V: Equivalents of the phrases ავი კაცი and ავი სიტყვა**

As mentioned above, the noun კაცი ‘man’ in NP1 is accompanied by the adjective ავი ‘evil’ in the source text. In English translations, the NP ავი კაცი ‘evil man’ is translated by formal equivalents (MW and LC). In some cases, the NP is replaced by N with a negative content (HH, HB, ShN and MB). In Spanish, the translator adopts the NP *un alma baja* ‘a poor soul’ by replacing the adjective ავი ‘evil’ with *baja* ‘poor’ (LM). The same strategy is used in the French translation by STs.: ავი კაცი ‘evil man’ > *une âme basse* ‘a poor soul’, but in this case the translator chooses as an equivalent for კაცი ‘man’ the connotative equivalent *âme* ‘soul’. In another translation, the noun phrase is very freely interpreted: *Verwegener* ‘temerarious, more daring’ (MP).

The second noun, სიტყვა ‘word’ in NP2, is also accompanied by the adjective ‘evil’ in the source language and is transferred in the same way in all the example translations, except for one in German which has *Hohn und Schmähung* ‘mockery and vituperation’ (HH). For ავი სიტყვა ‘evil word’ we thus find:

German: *böse Wörter* (MP), *böses Wort* (HB)

English: *evil word* (MW, LC)

French: *parole vénéneuse* (GB), *parole malsaine* (STs)

Russian: *Злое слово* (ShN)

Spanish: *palabra perversa* (GT), *palabra malsana* (LM), *palabras venenosas* (MB).

The selected NPs according to Nida’s “Formal and Dynamic” model of translation are given in Table VI:

	ავი კაცი	ავი სიტყვა
HH	D	D
MP	D	F
HB	F	F
MW	F	F
LC	F	F
GB	D	D
STs	D	D
ShN	F	F
GT	D	D
LM	D	D
MB	F	D

**Table VI: Formal and Dynamic equivalents of the phrases ავი კაცი and ავი სიტყვა**

Table VII shows the difference between equivalences of single words (word level) and NPs (phrase level), i.e., between კაცი vs. ავი კაცი and სიტყვა vs. ავი სიტყვა.

		კაცი	ავი კაცი	სიტყვა	ავი სიტყვა
GER	HH	D	D	D	D
	MP	D	D	F	F
	HB	F	F	F	F
ENG	MW	F	F	F	F
	LC	F	F	F	F
FR	GB	F	D	F	D
	STs	D	D	F	D
RU	ShN	F	F	F	F
ES	GT	F	D	F	D
	LM	D	D	F	D
	MB	F	F	F	D

**Table VII: Formal and Dynamic equivalents on the word and phrase level**

When selecting equivalents for the other two nouns, სული ‘soul’ and გული ‘heart’, four strategies can be observed:

- 1) In the target language, the two words are rendered identically by denotative equivalents, as in the two Spanish translations with სუელი > *alma* ‘soul’, გუელი > *corazón* ‘heart’ (GT and MB);
- 2) Only one of the two nouns is rendered in the target language:
  - a) სუელი > *Seele* ‘soul’ (HB), *soul* (LC) or
  - b) გუელი > *coeur* ‘heart’ (STs), გუელი > *corazón* ‘heart’ (LM);
- 3) In some translations, the two words are either rendered by one equivalent that represents a concrete noun (სუელი ‘soul’ + გუელი ‘heart’ > *жизнь* ‘life’) or they are rendered abstractly (სუელი ‘soul’ + გუელი ‘heart’ > *über alles* ‘more than anything else’);
- 4) The object of comparison (სუელი ‘soul’ + გუელი ‘heart’) is completely omitted, e.g. in the French translation by GB:

*L’homme méchant prise surtout une parole vénéneuse.*  
The wicked man takes especially venomous speech.

The next step would be to analyse the predication in the given sentence. In Georgian, this is represented by the verb ურჩევნია ‘prefer’. The equivalents of this verb in the translations mean ‘to love’, ‘to choose’ or ‘to prefer’ are illustrated in Table VIII.

	S	DO	V	O of comparison
ShR	ავსა კაცსა	ავი სიტყვა	ურჩევნია	სულსა, გულსა
HH	<i>Unhold</i> ‘fiend’	<i>Hohn und Schmähung</i> ‘mockery and vituperation’	<i>sind lieb</i> ‘to be dear, to love’	<i>(mehr als) Leib und Leben</i> ‘(more than) life and limb’
MP	<i>ein Verwegener</i> ‘a temerarious man’	<i>böse Wörter</i> ‘evil words’	<i>liebt</i> ‘to love’	<i>über alles</i> ‘above all’
HB	<i>der Böse</i> ‘the evil one’	<i>böses Wort</i> ‘evil word’	<i>liebt</i> ‘to love’	<i>(mehr als) die eigene Seele</i> ‘(more than ) one’s own soul’
MW	<i>an evil man</i>	<i>an evil word</i>	<i>loves</i>	<i>more than his soul or heart</i>
LC	<i>an evil man</i>	<i>an evil word</i>	<i>loves</i>	<i>more than he loves his soul</i>
GB	<i>l’homme méchant</i> ‘the wicked man’	<i>une parole vénéneuse</i> ‘a poisonous word’	<i>prise surtout</i> ‘to take especially’	—
STs	<i>une âme basse</i> ‘a poor soul’	<i>la parole malsaine</i> ‘the unhealthy word’	<i>choisit</i> ‘to choose’	<i>au lieu du coeur</i> ‘instead of the heart’
ShN	<i>лишь злодеи</i> ‘only/just a villain’	<i>злое слово</i> ‘evil word’	<i>почитаем</i> ‘to prefer’	<i>больше жизни</i> ‘above life’
GT	<i>un mal hombre</i> ‘a bad man’	<i>a una palabra perversa</i> ‘a wicked word’	<i>ama</i> ‘to love’	<i>más que a su alma y a su corazón</i> ‘more than his soul and heart’
LM	<i>Un alma baja</i> ‘a lowly/poor soul’	<i>la palabra malsana</i> ‘the unhealthy word’	<i>escoge</i> ‘to choose’	<i>en lugar del corazón</i> ‘instead of the heart’
MB	<i>el malvado</i> ‘an evil man’	<i>las palabras venenosas</i> ‘poisonous words’	<i>ama</i> ‘to love’	<i>más que el alma o el corazón</i> ‘more than the soul or the heart’

**Table VIII: Comparison of the sentence’s elements in the translations**

Two translations are near to identical in both the choice of equivalents for all words in the sentence and in the syntactic structure of the sentence (MW and LC):

MW *But an evil man loves an evil word more than his soul or heart.*

LC *But an evil man loves an evil word more than he loves his soul.*

The two following translations of the aphorism are also remarkably similar to each other but simultaneously differ much from the original version of the aphorism in the source language:

STs *Une âme basse, au lieu du coeur, choisit la parole malsaine.*

A poor soul, instead of the heart, chooses unhealthy speech.

LM *Un alma baja escoge la palabra malsana en lugar del corazón.*

A poor soul chooses the unhealthy word rather than the heart.

Some similarities in both cases are noteworthy:

- a) the noun კაცო ‘man’ is translated as ‘soul’
- b) the verb ურჩევნია ‘prefer’ is substituted by the verb ‘choose’ (STs: *choisit*, LM: *escoge*)
- c) the adjective ძვი ‘evil’ is transferred into the adjective ‘unhealthy’ (STs: *malsaine*, LM: *malsaina*)

Replacing the verb ‘prefer’ by the verb ‘choose’ has entailed certain morphosyntactic changes: Georgian ურჩევნია ‘prefer’ requires an object of comparison in addition to the DO (in our case, the nouns ‘soul’ and ‘heart’) while the verb ‘choose’ requires an object with an adposition like ‘instead’. Accordingly, the morphosyntactic structure of the sentence has been reorganised: instead of an object of comparison, an NP ‘instead of the heart’ is used as a prepositional object in *en lugar del corazón* (by LM) or as an embedded sentence construction (NP *au lieu du coeur* ‘instead of the heart’ by STs).

Because of the equivalent chosen, the object of comparison is omitted *in toto* in the French translation of GB. As mentioned above, the verb ‘prefer’ requires two objects (DO and object of comparison) and defines the priority of the agent. In translating this sentence, GB replaced the verb of the source text (‘prefer’) with a verb meaning ‘take’ (the person ‘takes’ a particularly poisonous speech). This verb change has also caused a change in the structure of the sentence: ‘take’ requires only a DO. Therefore, there is only one object in the translated text, and no object of comparison. Due to these changes, Rustaveli’s assertion is plainly simplified here. The aphorism has less expressiveness because, on the one hand, it cancels out the opposition between two objects, and on the other hand, it changes the meaning of the sentence: ‘evil people speak evil’ in this translation refers abstractly to a fact that it is the habit of evil people.

Table IX shows the translations of the sentence in terms of Formal and Dynamic equivalents.

	S	DO	V	O
ShR	ავსა კაცსა	ავი სიტყვა	ურჩევნია	სუფსა, გუფსა
HH	D	D	F	D
MP	D	F	D	D
HB	F	F	D	D
MW	F	F	D	F
LC	F	F	D	F(-)
GB	D	D	D	—
STs	D	D	D	F(-)
ShN	F	F	F	D
GT	D	D	D	F
LM	D	D	F	F(-)
MB	F	D	D	F

Table IX: Comparison of the sentence in the translations according formal and dynamic equivalents

The choice of lexical units as verb equivalents in the target language (operation on the lexical level) has led to morphosyntactic changes (agreement between verb and prime actants) which have an impact on the syntactic and pragmatic characteristics of the translated sentence in the target language as shown in Table X.

ShR	აგსა კაცსა LS (Dat.)	აჲო სიტყვა LDO (Nom.)	ურჩევნია V ( <i>verba sentiendi</i> )	სულსა, გულსა O (Dat)
HH	IO (Dat.)	S (Nom.)	V copula	O (Nom)
MP	S (Nom.)	DO (Acc.)	Vtr.	O (Acc.)
HB	S (Nom.)	DO (Acc.)	Vtr.	O (Acc.)
MW	S	DO	Vtr.	O
LC	S	DO	Vtr.	Synth.Const.
GB	S (Nom.)	DO (Acc.)	Vtr	—
STs	S (Nom.)	DO (Acc.)	Vtr	O (Nom)
ShN	S (Nom.)	DO (Acc.)	Vtr	O (Nom)
GT	S (Nom.)	DO (Acc.)	Vtr	O (Nom)
LM	S (Nom.)	DO (Acc.)	Vtr	O (Nom)
MB	S (Nom.)	DO (Acc.)	Vtr	O (Nom)

**Table X: Comparison of the sentence on the morphosyntactic level**

Let us now look at the information structure of the aphorism in the different translations.

**SOV** [Prep.O]

ShR	აგსა კაცსა	აჲო სიტყვა	ურჩევნია	სულსა, გულსა
-----	------------	------------	----------	--------------

**SVO** [Prep.O]

MW	<i>But an evil man</i>	<i>loves</i>	<i>an evil word</i>	<i>more than his soul or heart</i>
LC	<i>But an evil man</i>	<i>loves</i>	<i>an evil word</i>	<i>more than he loves his soul</i>
LM	<i>Un alma baja</i>	<i>escoge</i>	<i>la palabra malsana</i>	<i>en lugar del corazón</i>
MB	<i>El malvado</i>	<i>ama</i>	<i>las palabras venenosas</i>	<i>más que el alma o el corazón</i>

**S**[Prep.O]**VO**

STs	<i>Une âme basse</i>	<i>au lieu du coeur</i>	<i>choisit</i>	<i>la parole malsaine</i>
-----	----------------------	-------------------------	----------------	---------------------------

**SVO<sub>ind</sub>**[Prep.O]

HH	<i>Hohn und Schmähung</i>	<i>sind ... lieb</i>	<i>dem Unhold</i>	<i>mehr als Leib und Leben</i>
----	---------------------------	----------------------	-------------------	--------------------------------

**O**[Prep.O]**SV**

MP	<i>Böse Wörter</i>	<i>über alles</i>	<i>ein Verwegener</i>	<i>nur liebt</i>
----	--------------------	-------------------	-----------------------	------------------

**OVS** [Prep.O]

HB	<i>Böses Wort</i>	<i>liebt</i>	<i>der Böse</i>	<i>mehr als die eigene Seele</i>
----	-------------------	--------------	-----------------	----------------------------------

**O [Prep.O] VS**

ShN	<i>Злое слово</i>	<i>больше жизни</i>	<i>почитает</i>	<i>лишь злодей</i>
-----	-------------------	---------------------	-----------------	--------------------

Special cases consist in the addition (GT) or omission of information (GB):

GT	<i>Pero es cierto que</i>	<i>un mal hombre</i>	<i>ama</i>	<i>a una palabra perversa</i>	<i>más que a su alma y a su corazón</i>
GB	<i>L'homme méchant</i>		<i>prise</i>	<i>surtout une parole vénéneuse</i>	

Table XI illustrates the word order in the different translations.

	GEO	German			English		French		Rus.	Spanish		
	ShR	HH	MP	HB	MW	LC	GB	STs	ShN	GT	LM	MB
SOV	+											
SVO		+			+	+	+	+		+	+	+
OSV			+									
OVS				+					+			

**Table XI: Comparison of the word order in the source and target texts**

The choice of an SVO order is partly due to syntactic rules in the individual languages as in English, but partly also to the strategy of the translators.

**5. Analysis of aphorisms with a complex structure**

As it is generally known, Rustaveli uses aphorisms to offer wise guidance in a clear and intelligible manner. In the epic, wisdom is often dispersed across the stanzas. The poet defines the intended perspective and point of view (as thesis vs. antithesis) in the given stanza's opening two lines. A concluding advice with a broader concept can be found in the fourth line. The third line can be interpreted as a maxim, and a final sentence can serve as a reprimand, even though the final line is always given the decisive significance. Cf., e.g., the following stanza (0.24), with its analysis given in Table XII:

0.24 მიჯნურობა არის ტურფა, საცოდნელად ძნელი გვარი;  
 მიჯნურობა სხვა რამეა, არ სიძვისა დასადარი:  
 იგი სხვაა, სიძვა სხვაა, შუა უზის დიდი ზღვარი,  
 ნუვინ გაჰრევთ ერთმანერთსა, გესმას ჩემი ნაუბარი!

“Love is tender, it’s genus is difficult to recognise  
 Love is something else, it cannot be compared to fornication.  
 love is different, fornication is different, between them lies a big boundary,  
 Don't confuse it with each other, listen to my dispensation!”

<b>Thesis</b> <i>Definition</i>	მიჯნურობა არის ტურფა, საცოდნელად ძნელი გვარი “Love is beautiful, whose genus is difficult to know. <b>concept:</b> მიჯნურობა / love ( <b>C</b> ) <b>Predication:</b> <b>P1</b> (to be beautiful), <b>P2</b> (to be difficult to know)	Definition of concept <b>C</b> (მიჯნურობა) Predication: <b>P1</b> , <b>P2</b> , <b>C &gt; P1 + P2</b>
<b>Antithesis</b>	მიჯნურობა სხვა რამეა, არ სიძვისა დასადარი: Love is something else, it cannot be compared to fornication. <b>concepts:</b> მიჯნურობა / love ( <b>C</b> ) vs. სიძვა / fornication ( <b>c</b> )	Contradictory concepts <b>C</b> vs. <b>c</b> (სიძვა) Contradictory Predication: <b>P1</b> vs. <b>P2</b> <b>C ≠ c</b>
<b>Maxim</b>	იგი სხვაა, სიძვა სხვაა, შუა უხის დიდი ზღვარი It is different, fornication is different, between them lies a big boundary,	<b>C=X</b> vs. <b>c=Y X≠Y</b>
<b>Sentencia</b> <i>Conclusion</i>	ნუვინ გაჰრევთ ერთმანერთსა, გესმას ჩემი ნაუბარი! Don't confuse them with each other, listen to my dispensation!”	Admonition

**Table XII: Structure of stanza with concepts and predications**

The stanza ends with a (formula of) admonition: ნუვინ გაჰრევთ ერთმანერთსა ‘Don’t confuse them with each other’; in addition, this admonition is explicatively conveyed by the author: გესმას ჩემი ნაუბარი! ‘Listen to my dispensation!’.

Let us now look at the translation of this stanza in German:

**MP** *Des Midshnurempfindens Schöne wahrzunehmen fällt es schwer,  
Ungewohnt ist’s und erhaben – nicht der Unzucht gleich im Wert;  
Dem Gefühl dient – vor der Unzucht – eine strenge Grenz’ zur Wehr  
Man soll ja sie nicht verwechseln – schreibe ich mit Nachdruck her.*

“It is difficult to perceive the beauty of the *Mijnur* feeling,  
It is unusual and sublime – not equal in value to fornication;  
A strict limit serves as a defense against fornication  
One should not confuse them – I write emphatically.”

**HH** *Dieser hohen Minne Schönheit zu erfassen ist gar schwierig.  
Anders ist sie als die Wollust, ungemain und unbegierig,  
eine Welt trennt sie vom Laster, das nur eitel ist und irrig.  
Nicht sind beide zu verwechseln, glaubt mir, dies bezeuge hier ich.*

“It is very difficult to grasp this high love beauty.  
It is different from lust, uncommon and unhungry,  
a world separates them from vice, which is only vain and mistaken.  
The two cannot be confused, believe me, I am here to testify to this.”

**HB** *Schön ist die Liebe, gewiß, ein nur schwer zu begreifendes Etwas  
Liebe gleicht nicht dem niederen Huren, sie ist etwas andres,  
Eines ist Liebe, ein anderes Huren, durch Welten geschieden.  
Mengt sie nur nie durcheinander! Vernehmt ihr recht meine Worte?*

“Love is beautiful, certainly, something difficult to understand  
Love is not like the lowly whore, it is something different,  
One is love, another is whoring, separated by worlds.  
Just never mix them up! Do you hear my words correctly?”

The first line in the three German translations can be analysed in two parts as follows.



## Part 1:

ShR	მიჯნურობა არის ტურფა	Love is beautiful
MP	<i>Des Midshnurempfindens Schöne</i>	the beauty of the Midshnur feeling
HH	<i>Dieser hohen Minne Schönheit</i>	this high love beauty
HB	<i>Schön ist die Liebe</i>	Love is beautiful

## Part 2:

ShR	საცოდნელად ძნელი გვარი	(It is) difficult to recognise its genus
MP	<i>wahrzunehmen fällt es schwer</i>	It is difficult to perceive
HH	<i>zu erfassen ist gar schwierig</i>	It is very difficult to grasp
HB	<i>ein nur schwer zu begreifendes Etwas</i>	something difficult to understand

The structural analysis of the line according to concepts and predications is given in Table XIII.

ShR	Love is beautiful, difficult to recognise its genus	C1: concept of N1 ( <i>love</i> ) P1: predication of N1 ( <i>love</i> ) > COP ( <i>to be beautiful</i> ) C2: concept of N2 ( <i>a genus</i> ) P2: predication of N2 ( <i>a genus</i> ) > COP ( <i>to be difficult to recognise</i> )
MP	It is difficult to perceive the beauty of the Midshnur feeling	C1: concept: NP ( <i>the beauty of the Midshnur feeling</i> ) P1: predication of NP > COP ( <i>to be difficult to perceive</i> )
HH	It is very difficult to grasp this high love beauty	C1: Concept: NP ( <i>this high love beauty</i> ) P1: predication of NP > COP ( <i>to be very difficult to grasp</i> )
HB	Love is beautiful, certainly, something difficult to understand	C1: concept of N1 ( <i>love</i> ) P1: predication of N1 > COP ( <i>to be beautiful</i> ) C2: concept of N2 ( <i>something</i> ) P2: predication of N2 >COP ( <i>to be difficult to understand</i> )

**Table XIII: Contrastive analysis of the line in the German translations**

As Table XIV shows, HB's translation is very close to the source text in terms of both the adequacy of lexical equivalents and the structure of the sentence.

	Concepts	Predications
ShR	C1: concept of N1 ( <i>love</i> ) C2: concept of N2 ( <i>a genus</i> )	P1: predication of N1 > COP ( <i>to be beautiful</i> ) P2: predication of N2 > COP ( <i>to be difficult to recognise</i> )
HB	C1: concept of N1 ( <i>love</i> ) C2: concept of N2 ( <i>something</i> )	P1: predication of N1 > COP ( <i>to be beautiful</i> ) P2: predication of N2 >COP ( <i>to be difficult to understand</i> )

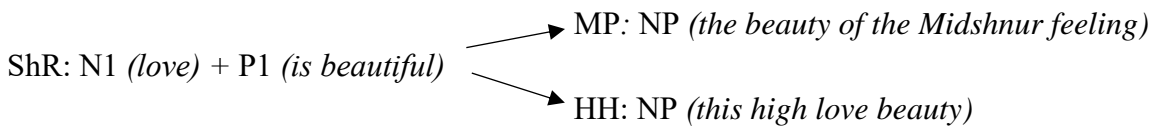
**Table XIV: Comparison of line structure and lexical equivalents in HB's translation**

The difference in concepts is that in C2, Rustaveli uses 'genus' in predication P2 while HB favours the indefinite pronoun 'something'. MP's and HH's translations show structural similarities, although they both differ from the structure of the source text as shown in Table XV.

	Concepts	Predications
ShR	C1: concept of N1 ( <i>love</i> ) C2: concept of N2 ( <i>a genus</i> )	P1: predication of N1 > COP ( <i>to be beautiful</i> ) P2: predication of N2 > COP ( <i>to be difficult to recognise</i> )
MP	C1: Concept: NP ( <i>the beauty of the Midshnur feeling</i> )	P1: predication of NP > COP ( <i>to be difficult to perceive</i> )
HH	C1: Concept: NP ( <i>this high love beauty</i> )	P1: predication of NP > COP ( <i>to be very difficult to grasp</i> )

**Table XV: Comparison of line structure and lexical equivalents in MP’s and HH’s translations**

The difference between the source text on the one hand and MP’s and HH’s translations on the other is that instead of a noun, the latter us a NP, which is a nominalisation of C1 + P1:



The structure of P1 is therefore the same in both translations:

- MP: P1: *predication of NP > COP (to be difficult to perceive)*
- HH: P1: *predication of NP > COP (to be very difficult to grasp)*

The examples considered here show that while translating, both translators use similar strategies. In our opinion, this type of structure analysis can be very helpful in researching the equivalence of aphorisms from a structural point of view.

Below, another stanza is introduced, with a different structure: it consists of three theses and a maxim.

29.718 გული კრულია კაცისა, ხარბი და გაუძღომელი,  
 გული ჟამ-ჟამად ყოველთა ჭირთა მთმო, ღხინთა მნდომელი,  
 გული - ბრმა, ურჩი ხედვისა, თვით ვერას ვერ გამზომელი,  
 ვერცა ჰპატრონობს სიკვდილი, ვერცა პატრონი რომელი!

“The heart of man is accursed, greedy and insatiable.  
 Sometimes the heart desiring joys endures all griefs.  
 The heart is blind, disobedient in seeing, not at all able to measure itself.  
 Nor death itself can master it, also nor an owner.”

The structural analysis of this stanza according to concepts and predications is given in Table XVI.

Thesis 1 Definition	გული კრულია კაცისა, ხარბი და გაუძღომელი, <i>The heart of man is accursed, greedy and insatiable</i>	C1: concept of NP1 ( <i>heart of man</i> ) P1: predication of NP1 > COP ( <i>to be accursed</i> ) P2: Predication of NP > COP ( <i>to be greedy</i> ) P3: Predication of NP > COP ( <i>to be insatiable</i> )
Sentiment: Negative	NP (NEU) > P1 (NEG) + P2 (NEG)+P3 (NEG)	

<p><b>Thesis 2</b> <i>Definition</i></p> <p>Sentiment: Negative</p>	<p>გული ჟამ-ჟამად ყოველთა ჭირთა მთმო, ლხინთა მნდომელი, <i>Sometimes the heart endures all griefs, desiring joys.</i></p> <p><i>NI (NEU) &gt; P4 (NEG) + P5 (POS)</i></p>	<p>C2: concept: of N1 (<i>heart</i>) [P4]: predication of N1 &gt; COP (<i>to endure griefs</i>) [P5] predication of N1 &gt; COP (<i>to desiring joys</i>)</p>
<p><b>Thesis 3</b> <i>Definition</i></p> <p>Sentiment: Negative</p>	<p>გული - ბრმა, ურჩი ხედვისა, თვით ვერას ვერ გამზომელი, <i>The heart is blind, disobedient in seeing, not at all able to measure itself.</i></p> <p><i>NI (NEU) &gt; P6 (NEG) + P7 (NEG) + P8 (NEG)</i></p>	<p>C2: concept of N1 (<i>heart</i>) [P6]: predication of N1 &gt; COP (<i>to be blind</i>) [P7]: predication of N1 &gt; COP (<i>to be disobedient in seeing</i>) [P8]: predication of N1 &gt; COP (<i>not to be not at all able to measure anything</i>)</p>
<p><b>Maxim</b> <i>Conclusion</i></p> <p>Sentiment: Negative</p>	<p>ვერცა ჰპატრონობს სიკვდილი, ვერცა პატრონი რომელი! <i>Nor death can master it, also nor an owner.</i></p> <p><i>N2 (NEU) + N3 (NEU) &gt; P9 (NEU)</i></p>	<p>C2: concept of [N1] (<i>heart</i>) C3: concept of N2 (<i>death</i>) C4: concept of NP2 (<i>an owner</i>) P9: predication of [N1] &gt; COP (<i>not to master</i>)</p>

**Table XVI: Description of the stanza's structure**

**Thesis 1:** negative definition of NP (*a man's heart*), which represents a set of negative qualities: accursed, greedy, insatiable.

**Thesis 2:** negative definition of N (*heart*), includes the set of properties: the heart [of a man] is changeable, from time to time he gives up a plague and looks for joy/happiness.

**Thesis 3:** negative definition of N (*heart*), the predication, which represents the set of negative properties: to be blind, to be disobedient in seeing, not to be not at all able to measure itself.

**Conclusion:** The heart is disobedient - Nor death can master it, also nor an owner.

In the following we compare the two English translations of this stanza with the source text.

**ShR** გული კრულია კაცისა, ხარბი და გაუძღომელი,  
გული ჟამ-ჟამად ყოველთა ჭირთა მთმო, ლხინთა მნდომელი,  
გული – ბრმა, ურჩი ხედვისა, თვით ვერას ვერ გამზომელი,  
ვერცა ჰპატრონობს სიკვდილი, ვერცა პატრონი რომელი!

**MW** Accursed is the heart of man, greedy, insatiable;  
sometimes the heart desiring joys endures all griefs;  
blind is the heart, perverse in seeing, not at all able to measure itself;  
no king, nor even death itself, can master it.

**LC** Accursed is the heart of man, greedy, insatiable, not smart;  
Sometimes the heart endures all grief because it seeks joy at the start.  
Blind is the heart, unable to distinguish the whole from the part.  
No king, not even death, can master the desires of a man's heart.

Table XVII illustrates the translation of the concepts of this stanza in the English translations.

	ShR	MW	LC	
1.	კაცის გული 'heart of man'	<i>heart of man</i>	<i>heart of man</i>	
2.	სიკვდილი 'death'	<i>death</i>	<i>death</i>	
3.	პატრონი 'master'	<i>king</i>	<i>king</i>	
4.	ჭირი 'grief'	<i>grief</i>	<i>grief</i>	
5.	ლხინი 'joy'	<i>joy</i>	<i>joy</i>	
6.	კრულია 'is accursed'	<i>is accursed</i>	<i>is accursed</i>	
7.	ხარბი 'greedy'	<i>greedy</i>	<i>greedy</i>	
8.	გაუძღომელი 'insatiable'	<i>insatiable</i>	<i>insatiable</i>	<i>not smart</i>
9.	ბრმა 'blind'	<i>blind</i>	<i>blind</i>	
10.	ურხი ხედვის 'disobedient in seeing'	<i>disobedient in seeing</i>	—	
11.	მომი 'compliant'	<i>to endures</i>	<i>endures</i>	
12.	მნდომელი 'desiring'	<i>desiring</i>	<i>because it seeks</i>	<i>at the start</i>
13.	თვით ვერას ვერ გამზომელი 'not able to measure anything itself'	<i>not at all able to measure itself</i>	<i>unable to distinguish</i>	<i>the whole from the part</i>
14.	ვერცა კპატრონობს 'cannot master'	<i>no one can master</i>	<i>no one can master</i>	<i>the desires of a man's heart</i>

**Table XVII: The translations of the concepts of this stanza in the English translations**

The comparison of the concepts of MW and LC shows the difference between their translation strategies: although both translations remain very close to the source text in terms of choice of equivalents, LC complements or interprets freely while translating.

ShR	გული	კრულია	კაცისა	ხარბი	გაუძღომელი	
	C1(Head)	P1 (Av+A)	C1(Att.)	P2 (A)	P3 (A)	
MW	Accursed	is	the heart of man	greedy	insatiable	
	P1 (A+ Av)		C1	P2 (A)	P3 (A)	
LC	Accursed	Is	the heart of man	greedy	insatiable	not smart
	P1 (A+Av)		C1	P2 (A)	P3 (A)	<i>P (Addition)</i>

We also see here how the structure within the sentence (which corresponds to the word order) changes. The concept C1 is an NP that has a postpositional structure (N+Attr.) and is additionally split by the copula (P1) in the source text. The copula inserted in NP gives more expressiveness to the characteristic of the heart, namely to be accursed.

გული	კრულია	კაცისა
C1(Head)	P1 (Av +A)	C1(Att.)

In both translations, C1 is not split (*the heart of man*) but it does not occur in the initial position, like in the source text; it is placed in the second position, after the copula (P1). Both translators choose a similar strategy and instead of the typical English structure, S(ubject) before V(erb), they change the word order in the sentence, which increases its meaningfulness:

Accursed	is	the heart of man
P1 (A+ Av)		C1

The translations of this verse by MW and LC are thus very similar both in translation technique and in the choice of equivalents, which indicates that LC has adopted MW’s translation. However, it is also evident that LC supplements and interprets the source text in order to further modify her translation as in the following examples.

- MW Accursed is the heart of man, greedy, insatiable;  
 LC Accursed is the heart of man, greedy, insatiable, **not smart**;
- MW Sometimes the heart desiring joys endures all griefs;  
 LC Sometimes the heart endures all grief **because** it seeks joy **at the start**.
- MW Blind is the heart, perverse in seeing, not at all able to measure itself;  
 LC Blind is the heart, **unable to distinguish the whole from the part**.
- MW No king, nor even death itself, can master it.  
 LC No king, not even death, can master **the desires of a man’s heart**.

The content structure of the stanza is retained in the translations but the concepts and predications do not match the source text in LC’s text. Here, the translator uses a free translation method comprising an addition or an interpretation, or both, as shown in Table XVIII.

	Structure of stanza: ShR (GE)	Structure of stanza: MW (ENG)	Structure of stanza. LC (ENG)	
Thesis 1	C1 (NP1), P1, P2, P3	C1 (NP1), P1, P2, P3	C1 (NP1), P1, P2, P3	Add.
Thesis 2	C2 (N1), P4, P5	C2 (N1), P4, P5	C2 (N1), P4, P5	Interp.
Thesis 3	C2 (N1), P6, P7, P8	C2 (N1), P6, P7, P8	C2 (N1), Cx (N3), Cy, Px	Add. Interp.
Concl.	C3 (N2), C4 (NP2), P9	C3 (N2), Cx (N3), C1 (Pron), P9	C3 (N2), Cx (N3), C1 (NP1), Cy (N4), P9	Add. Interp.

**Table XVIII: Comparison of content structure of the stanza in the English translations**

## Conclusions

The variety of languages into which Shota Rustaveli’s epic “The Knight in the Panther’s Skin” has been translated provides a unique opportunity for the creation of the multilingual parallel corpus ‘Rustaveli goes digital’, which is an outstanding resource for translation studies to examine the quality of translations, analyse translation methods, pinpoint translators’ strategies, and contrast them. Research into equivalence is a key factor in translation studies; it can be effectively used to evaluate the quality of translations. The translation quality of the aphoristic style in the translations is one of the challenges of the digital Rustvelology. In the present paper, we have discussed our research model on how to evaluate the translation quality of aphorisms.

The analyses of some aphorisms carried out here have shown that the equivalence of the aphorisms should be described and analysed on different structural levels (lexical, syntactic, pragmatic). The comparison of equivalence at the word and NP level has illustrated that there can be considerable differences in the evaluation and qualitative assessment of equivalents in the target language.

It is further important to analyse to what extent the content structure of the aphorisms is transferred. To this end, we carried out a structural analysis of the aphorisms and segmented them into the functional constituents (concepts and predications) to check whether they are adequately transferred into the target language and what changes they are subject to, namely whether the translations contain additions or interpretations, or both.

The model which we used to determine the degree of equivalence of aphorisms, is a conceptual model and requires verification of the feasibility of its use on larger material. In addition, the model has been applied to only a few languages at this time. In the future, we consider it necessary to analyse both the full corpus of aphorisms, on the one hand, and to increase the number of languages for analysis, on the other hand, in order to confirm the efficiency of our model in equivalence studies.

Only after a qualitative analysis of the equivalents of translated aphorisms will it be possible to conduct statistical analyses and to create a technological framework for equivalence studies in multilingual parallel corpora, which is one of the main challenges of the 21<sup>st</sup> century in translation studies.

## Abbreviations

### Grammatical terms:

A	adjective
Acc.	accusative
Add.	addition
Att.	attribute
Av	auxiliary verb
C / c	concept
D	dynamic equivalent
DAT	dative
F	formal equivalent
Foc	focus
Interp.	interpretation
LDO	logical direct object
LS	logical subject
N	noun
NEG	negative sentiment
NEU	neutral sentiment
NOM	nominative
NP	nominal phrase
Ocomp	object of comparison
Oind	indirect object
P	predication
Prep.O	prepositional object
PRES	present
S	subject
SG	singular

Synth.Const.	synthetic construction
V	verb
V sentiendi	verba sentiendi

### Languages:

ENG	English
ES	Spanish
FR	French
GER	German
RU	Russian

### Authors:

GB	Gaston Bouatchidzé
GT	Gustavo Alfredo de la Torre Botarro
HB	Herman Buddensieg
HH	Hugo Huppert
LC	Lyn Coffin
LM	Leonór Martínez
MB	MB – Maria Elvira Roca Barea
MP	Marie Prittwitz
MW	Marjorie Scott Wardrop
ShR	Shota Rustaveli
ShN	Shalva Nutsbidze
STs	Serge Tsouladzé

## References

- Bouatchidzé (1989): Gaston B., *Le chevalier à la peau de panthère*. Traduction revue par Philippe Dumaine et Bernard Outtier. Moscou: Radouga. Later edition: 2006.
- Buddensieg (1976): Hermann B., *Der Mann im Pantherfell. Altgeorgisches Epos*. Nachdichtung. Tbilisi: Sabtschota Sakartwelo. Later editions: 1989, 2003, 2005, 2011, 2012.
- Coffin (2015): Lyn C., Shota Rustaveli, *The Knight in the Panther Skin*. Tbilisi: Poezia Press.
- de la Torre Botarro (1962): Gustavo Alfredo d. I. T. B., Shota Rustaveli, *El Caballero de la Piel de Tigre. Epos heroico georgiano del siglo XII*. Traducción española. Santiago de Chile: [Edición Privada].
- Gabunia (2006): გიორგი გაბუნია, „შოთა რუსთაველის აფორიზმების ბიბლიოგრაფია“, შრომები, ტ. 1 (6), საქართველოს პარლამენტის ეროვნული ბიბლიოთეკა, გვ. 141–164.
- Gippert (2024): Jost G., “A zoological riddle from Medieval Georgia”, in Florian Mühlfried (ed.), *Languages and Cultures of the Caucasus: A Festschrift for Kevin Tuite*, Wiesbaden: Reichert, 85–103.
- Gobianidze (2024): Mariam G., “The Issue of Equivalence of Aphorisms in the English Translations of Shota Rustaveli’s ‘The Knight in the Panther’s Skin’ (based on one aphorism)”, *Millennium 2*, 84–101. DOI: <https://doi.org/10.62235/mln.2.2024.7994>.
- Huppert (2014): Hugo H., *Der Recke im Tigerfell. Ein Altgeorgisches Poem*. Deutsche Nachdichtung. Herausgegeben von Jost Gippert und Manana Tandaschwili. Wiesbaden: Reichert.
- Khintibidze (2018): ელგუჯა ხინტიბიძე, თანამედროვე რუსთაველოლოგიური კვლევებით კომენტირებული ვეფხისტყაოსანი, თბილისი [Elguja, Kh., *Contemporary Research in Rustvelology. Comments to ‘The Knight in the Panther’s Skin’*, Tbilisi]. ISBN 978-9941-9423-0-3.
- Martínez (2000): Leonór M., Rustaveli, *El Caballero de la Piel de Tigre*. Barcelona: Círculo de Lectores.
- Mellado Blanco (2010): Carmen M. B., “Die phraseologische Äquivalenz auf der System- und Textebene”, in: Jarmo Korhonen, Wolfgang Mieder, Elisabeth Piirainen, Rosa Piñel (eds.), *Phraseologie global – areal – regional*, Tübingen: Narr Francke Attempto, 277–284.
- (2015): Carmen M.B., “Parámetros específicos de equivalencia en las unidades fraseológicas (ejemplos del español y el alemán)”, *Revista de Filología* 33, 153–174.
- Mellado Blanco & Steyer (2018): Carmen M. B., Kathrin St., “Auf der Suche nach Äquivalenz, Lexikalisch geprägte Muster kontrastiv: Deutsch-Spanisch”, in Kathrin Steyer (ed.), *Sprachliche Verfestigung. Wortverbindungen, Muster, Phrasem-Konstruktionen* (Studien zur Deutschen Sprache, 79), Tübingen: Narr Francke Attempto, 265–284.
- Nida (1964): Eugene A. N., *Toward A Science of Translating*. Leiden / Boston: Brill.
- Nutsubidze (1940): Шалва Нуцубидзе., Шота Руставели, *Витязь в тигровой шкуре*. Поэма в стихах. Перевод с грузинского Шалва Нуцубидзе. Москва: Гослитиздат. Later editions: 1941, 1949, 1950, 1957, 1961, 1966, 1977, 1979, 1988, 2002.
- Prittwitz (2011): Marie P., *Der Ritter im Pantherfell. Ein altgeorgisches Epos*. Deutsche Nachdichtung. Herausgegeben von Steffi Chotiware-Jünger und Elgudsha Chintibidse. Aachen: Shaker.
- Roca Barea (2003): Maria Elvira R. B., Shota Rustaveli, *El Caballero de la Piel de Tigre*. Edición crítica y estudio preliminar. Málaga: Universidad de Málaga.
- Rustaveli (1957): შოთა რუსთაველი, ვეფხისტყაოსანი, რედ. ა. შანიძე, კ. კეკელიძე და ა. ბარამიძე, თბილისი: გამომცემლობა „სახელგამი“.

Tsouladzé (1964): Serge Ts., Chota Roustaveli, *Le chevalier à la peau de tigre*. Traduit du géorgien avec une introduction et des notes. Paris: Gallimard. Later editions: თბილისი: ლიტერატურა და ხელოვნება, 1966, 1989, 2013.

Wardrop (1912): Marjorie Scott W., Shot'ha Rust'haveli, *The Man in the Panther's Skin. A romantic epic*. London: Royal Asiatic Society. Later editions under the same title: 1966; 2001; 2004; 2005, 2007. Supplemented and revised edition by Elisabeth Orbelyani and Solomon Jordanishvili, with changed title: *The Knight in the Tiger's Skin*. Moscow: Society of Foreign Workers / New York: International Publishers, 1938; later editions: *The Knight in the Tiger Skin*: Moscow: Progress, 1977; Amsterdam: Fredonia, 2001; Tbilisi: Sharvadze, 2003; 2012; ΚΟΛΑΣΙ, 2012; 2013; *The Man in the Tiger Skin*: Tbilisi: Sharvadze 2010.

### **Online resources**

*Rustaveli goes digital*: <https://rustaveli-goes-digital.de/>

UNESCO: <https://en.unesco.org/silkroad/silk-road-themes/documentary-heritage/manuscript-collection-shota-rustavelis-poem-knight-panthers>



## ეკვივალენტობის კვლევისათვის მრავალენოვან პარალელურ კორპუსებში

მანანა თანდაშვილი (ფრანკფურტი), მარიამ ყამარაული (ჰამბურგი)

DOI: <https://doi.org/10.62235/dk.3.2024.8514>

[tandaschwili@em.uni-frankfurt.de](mailto:tandaschwili@em.uni-frankfurt.de) || ORCID: [0009-0005-7812-9124](https://orcid.org/0009-0005-7812-9124)

[mariam.kamarauli@uni-hamburg.de](mailto:mariam.kamarauli@uni-hamburg.de) || ORCID: [0009-0006-0404-4424](https://orcid.org/0009-0006-0404-4424)

შოთა რუსთაველის პოემა „ვეფხისტყაოსანი“ ქართული ლიტერატურის მწვერვალი, ხოლო აფორისტული სტილის თვალსაზრისით გამორჩეული შედეგია. ავტორის მსოფლმხედველობის კომპლექსურობა შუა საუკუნეებისა და რენესანსული აზროვნების შერწყმის შედეგს წარმოადგენს და აერთიანებს, ერთი მხრივ, ქრისტიანულ (და, ზოგადად, რელიგიურ) აზრებს ანტიკურ ფილოსოფიასთან (პლატონი, დიონისე არეოპაგელი, არისტოტელე) და, მეორე მხრივ, მითიურ და ტრანსცენდენტულ ხედვებს ანალიტიკურ აზროვნებასთან. მე-12 საუკუნეში შექმნილმა პოემამ განსაკუთრებული ადგილი მოიპოვა არა მხოლოდ ქართველთა კულტურულ მემკვიდრეობაში, არამედ მსოფლიო ლიტერატურის ისტორიაშიც - 2013 წლიდან პოემა შეტანილია იუნესკოს მსოფლიო კულტურული მემკვიდრეობის რეესტრში. დღეისათვის მოგვეპოვება „ვეფხისტყაოსნის“ 150-ზე მეტი თარგმანი მსოფლიოს 58 ენაზე, რაც უნიკალურ შესაძლებლობას იძლევა მრავალენოვანი პარალელური კორპუსის შესაქმნელად. ენათა სიმრავლე და სხვადასხვა დროს შესრულებული თარგმანების მრავალფეროვნება იშვიათ შესაძლებლობას იძლევა თარგმანმცოდნეობისათვის თარგმანების ხარისხის შესასწავლად, კერძოდ, თარგმანის მეთოდების გასაანალიზებლად, მთარგმნელთა მიერ გამოყენებული სტრატეგიების გამოსავლენად და შეპირისპირებითი ანალიზის ჩასატარებლად.

თარგმანმცოდნეობაში დღეისათვის არაერთი თეორია არის სამეცნიერო მიმოქცევაში, რომელიც ეკვივალენტობის საკითხს ეხება, ეკვივალენტების ორწევრა სისტემიდან (ნაიდას კლასიფიკაცია) ხუთწევრა სისტემის (კოლერის კლასიფიკაცია) ჩათვლით. თუმცა პრობლემური რჩება საკითხი, როგორ შევაფასოთ თარგმანების ხარისხი მეტაფორული სტილის, განსაკუთრებით აფორიზმების შემთხვევაში? ცნობილია, რომ წყაროსა და სამიზნე ტექსტებს შორის ეკვივალენტობის ხარისხის დადგენა ლექსიკურ დონეზე<sup>11</sup> არ არის საკმარისი ლიტერატურული ნაწარმოებების, განსაკუთრებით პოეტური ძეგლების თარგმანების ხარისხობრივი შეფასებისთვის. საჭიროა ეკვივალენტობის შეფასება სხვადასხვა ენობრივ დონეზე (მაგ., მორფოლოგიურ, მორფო-სინტაქსურ, სინტაქსურ, სემანტიკურ, პრაგმატურ დონეებზე). თუმცა, როგორც ჩვენ მიერ შექმნილ მრავალენოვან კორპუსში „რუსთაველი ციფრულ ეპოქაში“ (*Rustaveli goes digital*), რომელიც დღეისათვის 38 თარგმანს აერთიანებს 25 ენაზე,

<sup>11</sup> პარალელურ კორპუსებში ამ ოპერაციის განსახორციელებლად გამოყენებულია ალგორითმების მეთოდი.

„ვეფხისტყაოსნის“ თარგმანებში ეკვივალენტობის კვლევამ მიგვიყვანა იმ დასკვნამდე, რომ ახალი, სტრუქტურული ანალიზის ჩარჩო უნდა შეიქმნას პოეტურ ნაწარმოებებში ეკვივალენტობის საკვლევად.

წინამდებარე ნაშრომში განხილულია თარგმანის ხარისხის, კერძოდ ეკვივალენტობის კვლევის თეორიული საფუძვლები და კვლევის მეთოდები, და წარმოდგენილია სტატიის ავტორთა მიერ შემუშავებული კონცეპტუალური მოდელი, რომელიც შეიძლება გამოვიყენოთ აფორიზმების ეკვივალენტობის შეფასებისას პარალელურ კორპუსში, კერძოდ, „ვეფხისტყაოსნის“ თარგმანების მრავალენოვან პარალელურ კორპუსში.

„ვეფხისტყაოსნის“ შემთხვევაში დამატებით სირთულეს ქმნის პოემაში აფორიზმების როგორც კვანტიტატიური, ისე კვალიტატიური ასპექტი. კერძოდ, ის ფაქტი, რომ აფორიზმების რაოდენობა, რომელიც აფორიზმების 30-მდე ერთენოვანი ან მრავალენოვანი კრებულების სახით არის გამოცემული, მერყეობს 24-დან 234-მდე, რაც იმაზე მიუთითებს, რომ „ვეფხისტყაოსნის“ აფორიზმების კვალიფიკაციის საკითხი ჯერ კიდევ საკამათოა. გარდა ამისა, იმის მიუხედავად, რომ აფორიზმების განსაზღვრების ერთ-ერთ ძირითად პოსტულატს წარმოადგენს დებულება, რომ აფორიზმი არის მოკლე, სხარტად გადმოცემული სიბრძნე, „ვეფხისტყაოსანში“ არაერთი შემთხვევა გვაქვს, როდესაც აფორიზმად მიჩნეულია არა მხოლოდ მოკლედ, ერთი წინადადებით გადმოცემული ბრძნული აზრი, არამედ ორ-, სამ- და ოთხსტრიქონიანი კონსტრუქციებიც კი (ანუ მთელი სტროფი).

სტატიაში განხილულია სხვადასხვა სახის - როგორც მარტივი, ისე რთული კონსტრუქციებით გადმოცემული აფორიზმების სტრუქტურული ანალიზი და თარგმანში აფორიზმების ეკვივალენტობის შეფასების კონცეპტუალური მოდელი.

როგორც წყარო და მიზანენაში აფორიზმების შედარებით-შეპირისპირებითმა კვლევამ გვიჩვენა, მარტივი სტრუქტურის მქონე აფორიზმების ეკვივალენტობის დადგენა სიტყვის დონეზე ვერ იძლევა ეკვივალენტობის ადეკვატურ სურათს. აფორიზმის „ავსა კაცსა ავი სიტყვა ურჩევნია სულსა, გულსა“ განხილვის მაგალითზე ნაჩვენებია, რომ ეკვივალენტობის დადგენა მხოლოდ ფრაზულ დონეზე იძლევა ადეკვატურ შედეგს. აქ განხილული ემპირიული მასალა მოიცავს 5 ენის 11 თარგმანს:

		S	DO	V	O
	შ.რ.	ავსა კაცსა	ავი სიტყვა	ურჩევნია	სულსა, გულსა
გერმ.	HH	D	D	F	D
	MP	D	F	D	D
	HB	F	F	D	D
ინგ.	MW	F	F	D	F
	LC	F	F	D	F(-)
ფრ.	GB	D	D	D	—
	STs	D	D	D	F(-)
რუს.	ShN	F	F	F	D
ესპ.	GT	D	D	D	F
	LM	D	D	F	F(-)
	MB	F	D	D	F

ტაბულა 1: აფორიზმის შედარება თარგმანებში ფორმალური და დინამიკური ეკვივალენტობის თვალსაზრისით

აღნიშნული აფორიზმი ასევე გავანალიზეთ ინფორმაციული სტრუქტურის თვალსაზრისით. შედეგების შედეგები მოცემულია ტაბულა 2-ში:

	ქართ.	გერმანული			ინგლისური		ფრანგული		რუს.	ესპანური		
	შ.რ.	HH	MP	HB	MW	LC	GB	STs	ShN	GT	LM	MB
SOV	+											
SVO		+			+	+	+	+		+	+	+
OSV			+									
OVS				+					+			

ტაბულა 2: აფორიზმის ინფორმაციული სტრუქტურის შედარება თარგმანებში

ენათა უმრავლესობაში დადასტურდა SVO სტრუქტურა, რაც ნაწილობრივ ენათა შინაგანი სტრუქტურის, კერძოდ სინტაქსური წესების რეალიზების აუცილებლობით არის განპირობებული (როგორც მაგ., ინგლისურში), თუმცა, ზოგიერთ შემთხვევაში თავად მთარგმნელთა ნებაყოფლობით სტრატეგიაზეც მიუთითებს. ყველაზე იშვიათად დადასტურებული OVS სტრუქტურა ჰერმან ბუდენზიგის გერმანულ თარგმანში არ ასახავს გერმანული ენის სინტაქსის კანონიკურ მოდელს, იგი აშკარად მთარგმნელის ინდივიდუალური გადაწყვეტილების შედეგია, აფორიზმში პრაგმატულ დონეზე მარკირებული ინფორმაციული სტრუქტურის თავისებურება წყაროენაში (ზმნის წინ მდგარი ლოგიკური ობიექტი, რომელიც სახელური ფრაზით *ავი სიტყვა* არის გადმოცემული, ქართულ ტექსტში ზმნის წინ დგას და ფოკუსირებულია) ადეკვატურად გადმოიტანოს გერმანულ თარგმანში - განათავსოს იგი წინადადების მარცხენა პერიფერიაში, ინიციალურ პოზიციაში და შეუნარჩუნოს ფოკუსი.

რთული სტრუქტურის შემთხვევაში კიდევ უფრო რთულდება აფორიზმების ეკვივალენტობის კვლევა, ვინაიდან აქ ხშირად მთლიანი სტროფის ფარგლებში გვიხდება ეკვივალენტობის ხარისხის დადგენა. აფორიზმის სტრუქტურის სირთულიდან გამომდინარე, შესაბამისად, შევიმუშავეთ შლადი კომპონენტური ანალიზის მეთოდი, რომელიც ორ ეტაპად ხორციელდება: 1. სტროფები დაყვავით სტრიქონებად და დაგადგინეთ მათი ლოგიკურ-ფუნქციური მნიშვნელობა აფორიზმში (მაგ., თეზა, ანტითეზა, მაქსიმა, სენტენცია) და 2. სტრიქონები დაყვავით კონცეპტებად და პრედიკაციებად და ცალ-ცალკე შევადარეთ თარგმანში შესაბამისად დამუშავებულ სტროფებს შინაარსობრივი ეკვივალენტობის შეფასების მიზნით. წყაროენასა და მიზანენაში კონკრეტული აფორიზმების ანალიზის შედეგად გამოვლინდა ის სტრუქტურული სხვაობები, რომელსაც განიცდიან რთული სტრუქტურის აფორიზმები მიზანენაში გადატანის დროს. საანალიზო მასალაში დადასტურდა როგორც სტრუქტურული კლების, ისე მატების ან ინტერპრეტაციების შემთხვევები, ან ორივე ერთად.

მაგალითად:

- გული კრულია კაცისა, ხარბი და გაუძღომელი,
- გული ჟამ-ჟამად ყოველთა ჭირთა მთმო, ლხინთა მნდომელი,
- გული – ბრმა, ურჩი ხედვისა, თვით ვერას ვერ გამზომელი,
- ვერცა ჰპატრონობს სიკვდილი, ვერცა პატრონი რომელი!

შდრ. მარჯორი უორდროპისა (MW) და ლინ კოფინის (LC) თარგმანები სტრიქონების მიხედვით:

- MW Accursed is the heart of man, greedy, insatiable;  
 LC Accursed is the heart of man, greedy, insatiable, **not smart**;
- MW Sometimes the heart desiring joys endures all griefs;  
 LC Sometimes the heart endures all grief **because** it seeks joy **at the start**.
- MW Blind is the heart, perverse in seeing, not at all able to measure itself;  
 LC Blind is the heart, **unable to distinguish the whole from the part**.
- MW No king, nor even death itself, can master it.  
 LC No king, not even death, can master **the desires of a man's heart**.

ტაბულა 3-ში მოცემულია აფორიზმის შლადი კომპონენტური ანალიზის მეთოდის გამოყენების შედეგები:

	სტროფის სტრუქტურა ქართულში	სტროფის სტრუქტურა ინგლისურში (უორდროპი)	სტროფის სტრუქტურა ინგლისურში (კოფინი)	
თეზა 1	C1 (NP1), P1, P2, P3	C1 (NP1), P1, P2, P3	C1 (NP1), P1, P2, P3	მატება
თეზა 2	C2 (N1), P4, P5	C2 (N1), P4, P5	C2 (N1), P4, P5	ინტერპ.
თეზა 3	C2 (N1), P6, P7, P8	C2 (N1), P6, P7, P8	C2 (N1), Cx (N3), Cy, Px	მატება, ინტერპ.
დასკვნა	C3 (N2), C4 (NP2), P9	C3 (N2), Cx (N3), C1 (Pron), P9	C3 (N2), Cx (N3), C1 (NP1), Cy (N4), P9	მატება, ინტერპ.

ტაბულა 3: აფორიზმის კომპონენტური შედარება თარგმანებში

თარგმანებში აფორიზმების ეკვივალენტობის შეფასებისათვის გამოყენებული მეთოდები, რომელიც ჩვენ მიერ შემოთავაზებული მოდელის ფარგლებში შევიმუშავეთ, ბუნებრივია შემდგომ დახვეწა-განვითარებას მოითხოვს. სასურველია მოხდეს მოდელის ევალუაცია უფრო მეტ მასალაზე და მეტ ენაზე, რათა დადგინდეს მისი გამოყენების ეფექტურობის ხარისხი. დადებითი შედეგების მიღების შემთხვევაში შეიქმნება ეკვივალენტობის ხარისხის დადგენის ტექნოლოგიური მოდელი, რომელიც საშუალებას მოგვცემს ავტომატურად, დიგიტალური მეთოდების გამოყენებით ვიკვლიოთ ეკვივალენტობა მრავალენოვან პარალელურ კორპუსში.

Paul's First Epistle to the Corinthians ch. 13:  
A Comparison of Georgian and Abkhaz Translations  
(Taking into account the Greek Original)

George Hewitt (London)

DOI: <https://doi.org/10.62235/dk.3.2024.8516>

[gh2@soas.ac.uk](mailto:gh2@soas.ac.uk) || ORCID: [0000-0002-7330-4107](https://orcid.org/0000-0002-7330-4107)

**Abstract:** The Bible is universally recognised to be the most translated book in the world. The whole Bible (Old & New Testaments) is, unsurprisingly, available in Georgian, which has a centuries-old literary tradition. In Abkhaz, however, only one version of the whole New Testament (NT) currently exists, and that is the privately published translation (from Russian) by the late writer Mushni Lasuria (2004). I spent several months a few years ago on behalf of *The Institute for Bible Translation* (once based in Stockholm but now in Moscow) going through Lasuria's translation of the four Gospels (together with a different translation, also done from Russian, by Zaira Khiba) and suggesting changes to align texts with the Greek original – Khiba's version of the Gospels is available online.<sup>1</sup> I therefore had the idea that it might prove to be an interesting and appealing exercise to take another extremely well-known part of the NT, namely Paul's discourse on the theme of love, and examine how this has been treated by Lasuria and Georgian translators (in both Old and Modern Georgian), contrasting the results with the ultimate Greek source. Given that the 13<sup>th</sup> chapter of Paul's First Epistle to the Corinthians is (one might say) universally regarded as a description of love, it might come as a surprise that not all English translations actually include this word, preferring *charity* instead. We shall touch on the reason why this choice caused such a furious controversy in 16<sup>th</sup>-century England.

**Keywords:** Abkhaz, Georgian, Greek, Latin, Russian; Bible, New Testament, St. Paul, Corinthians, St. Jerome, Vulgate, Tyndale, Institute for Bible Translation, Patriarchate; Charity, Love; Lasuria

## Introduction

Having spent several months a few years ago on behalf of The Institute for Bible Translation checking Mushni Lasuria's Abkhaz translation (done from Russian) of the four Gospels (together with a different translation, also from Russian, by Zaira Khiba) and suggesting changes where I felt it necessary in order to align the texts with the Greek original, I thought readers might be interested in language-issues raised by an examination of Lasuria's and the various Georgian translations of another extremely well-known part of the New Testament (NT). This is St. Paul's discourse on the theme of love in chapter 13 of his First Epistle to the Corinthians. The comparison offered below takes into account the ultimate Greek source and, of course, translations in both Modern and Old Georgian available to me in my personal collection. Before listing the Georgian translations consulted, I wish to stress that I am not interested either in theological issues or in making any value-judgement about which translation is the best; I am solely concerned with the linguistic question of how the text is rendered, given the resources of the two unrelated but neighbouring languages.

---

<sup>1</sup> <http://georgehewitt.net/articles/miscellaneous/316-the-four-gospels-in-abkhaz-translated-from-russian-by-zaira-kiazimovna-khiba>.

## 1. The Relevant Bible Translations in Georgian

Lasuria's translation was, as far as I know, the result of his own initiative and was privately published in 2004. As for Modern Georgian, a glance at the list of works consulted reveals that the text I have chosen is contained in five publications, three of which present the whole Bible, whilst the other two offer the New Testament and Psalms. Of these five, four were printed in Stockholm under the aegis of *The Institute for Bible Translation*. The 2002 edition explains the relationships between the cited versions. I quote:<sup>2</sup>

In 1989–1990 *The Institute for Bible Translation* brought out an experimental translation of the Bible in four volumes. This work too was accomplished far from Georgia, and so it required editing, which was effected by Georgian scholars; indeed, some books they translated anew. Today there already exist two new redactions of the Bible in the Georgian language (one, prepared by the Patriarchate of Georgia and published in 1989, and the second, the current translation completed at *The Institute for Bible Translation*).

Further details are presented in the Introduction to the 1980/1991 edition of the NT with Psalms. It is explained that the initial work was carried out over five years (1974–79) by a translator whose initials were G. Ts'. We then read:<sup>3</sup>

In 1982 we put out the renewed text of the NT, to which we appended the Psalms earlier translated from Ancient Hebrew. Thereafter, our translation was published several more times, almost unaltered. But now we humbly offer the reworked redaction of the said translation, which was carried out by Prof. Z. K'ik'nadze and his pupil M. Songhulashvili with the prayers and blessing of Ilia II, Catholicos-Patriarch of the whole of Georgia.

Given this background-information, it should come as no surprise to learn that, whilst not identical, the versions issued under the aegis of *The Institute for Bible Translation* (hereafter IBT) are very close to one another, but the publication from the Patriarchate (hereafter PV) is clearly divergent, as we shall see. For ease of reference, I have scanned (a) the Epistle's 13<sup>th</sup> chapter from *Bagster's Critical New Testament*, which combines the Greek original along with both English interlinear glosses and a more literary English rendition (Fig. 1),<sup>4</sup> (b) the Georgian text printed in the Patriarchate's large-format volume (Fig. 2),<sup>5</sup> and (c) Lasuria's Abkhaz text (Fig. 3) and provided the latter with interlinear English glosses (Appendix 1) and my English translation (Appendix 2).

We do not have the space to discuss every point of divergence, and so I shall start with selected observations on the Georgian versions in terms of (i) vocabulary, (ii) verb-mood, and (iii) other features.

### 1.2. Vocabulary

Verse 1 alone provides several examples of lexical diversity, starting with PV's very first word, 'of men' (in the phrase 'languages of men') being rendered by the old Genitive plural (in *-ta*) of the root *k'ats-* (კატსოძ), whereas all the IBT versions prefer the root *adamian-*, one (namely

<sup>2</sup> *Biblia* 2002: unnumbered page (Foreword, translation G.H.).

<sup>3</sup> *NT* 1980/1991: unnumbered page (Foreword, translation G.H.).

<sup>4</sup> *NT* n.d.: 692–693.

<sup>5</sup> *Bible* 1989: 1143.



NT 1982) using the modern plural (-*eb-isa* = დღამიანებისა). For the phrase ‘the echoing sound of brass’ we find the word for ‘brass’ (რკადლი) with PV’s postposed adjective *mzhghriale* = მოჟღერიალჳ (which reflects the Greek word-order) but with preposed *zhgharuna* = ჳღარუნა elsewhere, whilst for the sound of a cymbal (apparently there is a choice between either *ts’intis’ili* or *ts’intis’ila*) the offerings are PV *ts’k’riala* = ჳკრიალა vs *chxaruna* = ჩხარუნა elsewhere, though NT 1982 again diverges with *mzhgheri* = მოჟღერი.

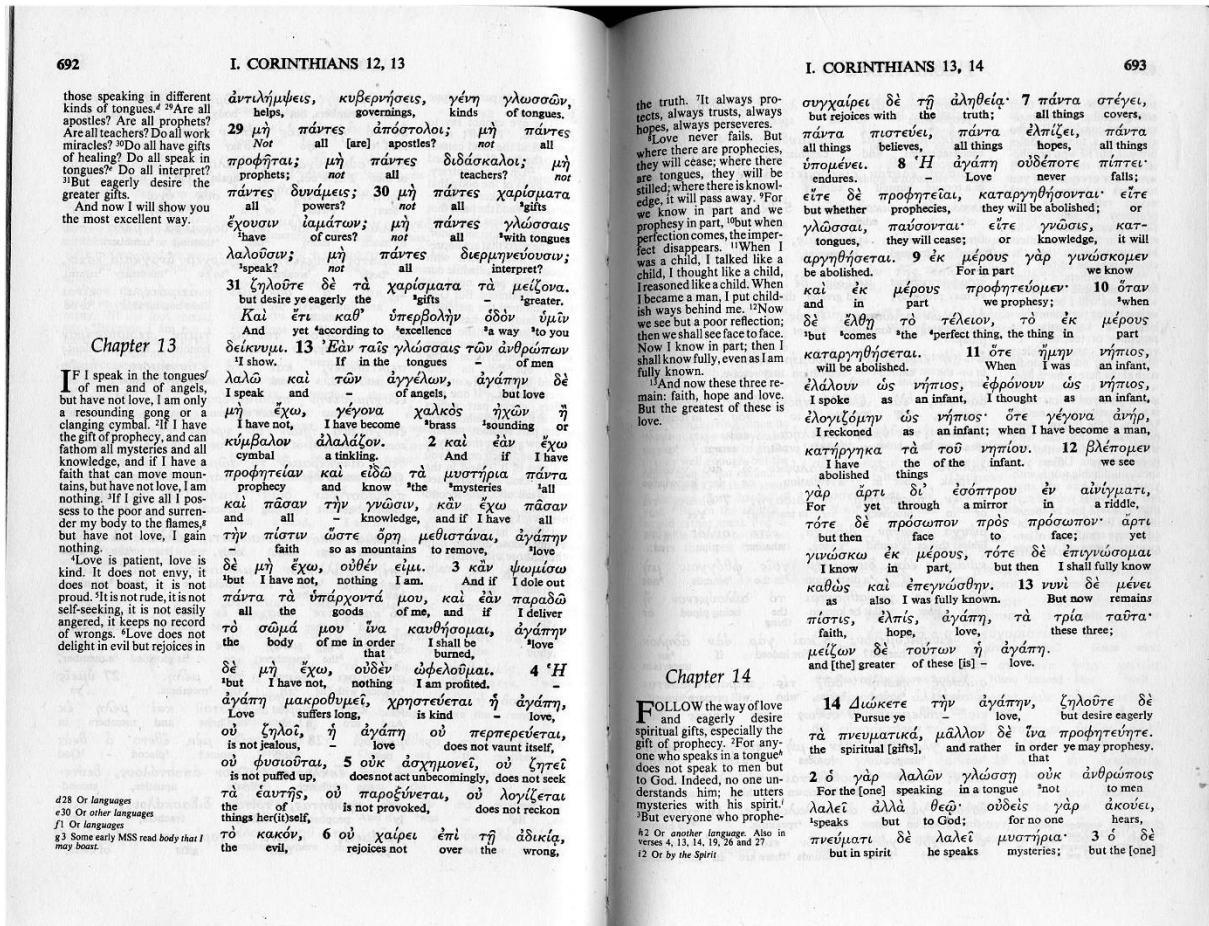


Fig. 1: Paul's First Epistle to the Corinthians ch. 13: Greek text along with both English interlinear glosses and a more literary English rendition

Of course, where (near-)synonyms exist, one is free to choose according to one's preferences, but a translator might like to remain as close as possible to the original. Take verse 5, where Greek captures in a single verb the meaning ‘behave unbecomingly’. And so the sequence ‘[love] does not behave unbecomingly’ corresponds to a single verb in PV’s *ar uk’eturobs* = არ უკეთურობს, whilst all the IBT versions expand to give *ar schadis uts’esobas* = არ სხადის უწესობას ‘does not commit impropriety’. PV is similarly closest to Greek’s single verb ‘I-am-profitd’ in verse 3 by employing the Stative verb (*aras*) *margia* = არას მარჯია ‘there-is-profit-to-me (in nothing)’, whereas three of the IBTs expand to produce *araperi sargebeli makvs* = არავეერი სარგებელი მაქვს ‘I have nothing of advantage’. But what do we find in NT 2002? Whilst the same root (viz. -*rg-* = -რგ-) is used as in PV, there seems to be a change in the meaning, because *araprad vevargebi* = არაფრად ვეპარვები looks to me to signify

თავი მცხამეტი

1. კაცთა და ანგელოზთა ენებზეც რომ ვმეტყველებდე, სიყვარული თუ არა მაქვს, მხოლოდ რვალი ვარ მოყვრილი, მხოლოდ წკრიალა წინწილი.
2. წინასწარმეტყველების მადლი რომ მქონდეს, ვიცოდე ყველა საიდუმლო, და მქონდეს მთელი რწმენა, ისე რომ მთების დაძვრაც შემიძლოს, სიყვარული თუ არა მაქვს, არარა ვარ.
3. მთელი ჩემი ქონება რომ გავიღო გლახათათვის, და დასაწვავად მივცე ჩემი სხეული, სიყვარული თუ არა მაქვს, არას მარგია.
4. სიყვარული სულგრძელია და ტკბილი; სიყვარულს არ შურს; სიყვარული არ ქედმაღლობს, არ ზვობს;
5. არ უყეთურობს, არ ეძებს თავისას, არ მრისხანებს, არ განიზრახავს ბოროტს;
6. არ შეჰხარის უსამართლობას, არამედ ჭეშმარიტებით ხარობს;
7. ყველაფერს იფარავს, ყველაფერი სწამს, ყველაფრის იმედი აქვს, ყველაფერს ითმენს.
8. სიყვარული არასოდეს არ გადავა, თუმცა წინასწარმეტყველებანი განქარდებიან, ენები დადუმდებიან, და უქმი გახდება ცოდნა.
9. რადგან ჩვენ ნაწილობრივ ვიცით, ნაწილობრივ კი ვწინასწარმეტყველებთ.
10. ხოლო როდესაც მოიწევა სრულქმნილება, მაშინ განქარდება ნაწილობრივი.
11. როდესაც ვიყავი ყრმა, ვიცოდი, როგორც ყრმა, ვფიქრობდი, როგორც ყრმა, ვმსჯელობდი, როგორც ყრმა; ხოლო როდესაც დავეკაცდი, ზურგი ვაქციე ყოველივე ყრმობისდროინდელს.
12. რადგან ახლა ბუნდოვნად ვხედავთ, როგორც სარკეში, ხოლო მაშინ პირისპირ ვიხილავთ; ახლა ნაწილობრივ ვიცით, ხოლო მაშინ შევიცნობ, როგორც თავად შევიმეტნები.
13. ჯერჯერობით კი ეს სამია: სარწმუნოება, სისოება და სიყვარული; ხოლო ამათში უმეტესი სიყვარულია.

თავი 13. (2) მათ. 7,22. (3) მათ. 6,1. (4) იგავ. 10,12. 1 პეტრ. 4,8. ფილიპ. 2,3. (5) რომ. 15,1. (7) გალ. 6,2. (9) 2 კორ. 5,7. (12) ფილიპ. 3,12. 2 ტიმ. 2,19.

1143

Fig. 2: Georgian text of I Corinthians ch. 13 printed in the Patriarchate’s large-format volume

‘I shall be of no use at all (sc. to anyone/anything else)’.<sup>6</sup> Verse 11 illustrates the reverse pattern, for the Greek has a verb plus noun-complement, as does English ‘(I) became a man’, which is mirrored in the IBT versions’ (*mama*)*k’atsi gavxdi*, but in this case PV employs just the simple verb *davk’atsdi* = დავეკაცდი.

1.3. Verb-mood

Chapter 13 starts with a series of conditional clauses formed in Greek with the particle *an* in combination with the subjunctive mood, which together serve to distance the event from the present. This is reflected precisely only in the very first clause in PV, namely: ... *rom vmet’q’velebde* = რომ ვმეტყველებდე ‘... if I were speaking’, whereas the IBT versions present the action as applicable to the here and now, viz. ... *tu vlap’arak’ob* = თუ ვლაპარაკობ ‘... if I am speaking’, and even PV quickly shifts to the Present Indicative in the adjoining conditional, viz. *siq’varuli tu ara makvs* = სიყვარული თუ არა მაქვს ‘... if I have no love’ as against the Greek’s remoter ‘but if I had not love’ (rather than ‘if I have not love’); the Old Georgian manuscripts have the subjunctive in both clauses, viz. *ghatu... vit’q’odi* [Subjunctive] *da siq’uaruli tu ara makundes //...xolo siq’uaruli ara makundes* = დათუ... ვიტყვოდი და სიყვარული თუ არა მაქუნდეს //...ხოლო სიყვარული არა მაქუნდეს.

<sup>6</sup> There is no doubt about the passive reading of οὐδὲν ὠφελῶμαι ‘I am profited in no way’. Cf. Sophocles *Antigone* l.550, where Ismene asks her sister Antigone this question: τί ταῦτ’ ἀνὰξ μ’, οὐδὲν ὠφελουμένη; ‘Why are you causing me such distress, being (yourself) in no way profited?’. The Latin Vulgate has: *nihil mihi prodest*.



#### 1.4. Other Features

The Greek text before us has St. Paul ascribing to himself in verse 11 three actions when he was a child: 'I talked, thought, reasoned (as a child)'. Three corresponding verbs are found only in PV and NT 1982, where the verbs are respectively: (PV) *vit'q'odi*, *vpikrobdi*, *vmsdzhelobdi* = ვიტყოდი, ვფიქრობდი, ვმსჯელობდი vs (NT 1982) *vlap'arak'obdi*, *vazrovnebd*, *vmsdzhelobdi* = ვლაპარაკობდი, ვაზროვნებდი, ვმსჯელობდი. The remaining three IBT versions have only the one verb *vmsdzhelobdi* = ვმსჯელობდი. Why is this? I imagine that differing Greek versions must underlie the disparity. If we look at the Old Georgian rendition according to the 1963 publication,<sup>7</sup> we find just the one verb *vit'q'ode* = ვიტყოდე [Indicative!] 'I used to speak', which matches the CD redaction of the Pauline Epistles, whilst in the AB redaction we have the two verbs *vit'q'ode* = ვიტყოდე [Indicative!] and *gulixma-vhq'opd* = გულისხმა-გვეყოფდ 'I used to speak and understand/reason'.<sup>8</sup> What follows in all three texts is: *sheratsxil viq'av vitartsa q'rmaj* = შერაცხილ ვიყავ ვითარცა ყრმაჲ 'I was counted as a child'. I think the explanation must be that the early translators regarded the Greek Imperfect ἐλογιζόμην of what was a so-called deponent verb (which is to say that it was passive in form but active in meaning) as having here the actual passive sense of 'I used to be counted/reckoned', whereas the translators who produced a third verb which can be translated as 'I used to count/reckon/reason', namely actively.<sup>9</sup>

In summation, it is plainly reasonable to conclude that scrutiny of the Georgian versions reveals a pretty close adherence to the Greek source(s). Variations between the translations are relatively minor and can mostly be attributed to the preferences of translators given a range of (near-)synonyms in the stock of nouns, verbs, adjectives, etc. Native speakers are then best placed to judge if the appropriate choices of the translator(s) accord with their own preferences.

## 2. Critical Analysis of Lasuria's Bible Translation in Abkhaz

What do we find when we turn to the only currently available translation in Abkhaz? Even without any knowledge of Abkhaz or the phonetic values of its letters in the text copied in Appendix 1, it must be obvious from the interlinear word- (though not morpheme-)glosses that matters are rather different from what we have seen so far.

a) The sense of distance from the present is nicely captured by the use of the subjunctive mood (in *-a:jt'* = *-aaum* in the sixth word of Verse 1 and later), but a problem arises at the end of the verse. St. Paul compares his speech to two sounds, but Lasuria merges them to produce a single analogy, namely, 'My voice resembles the dull sound that emanates from bronze plates clashing against each other'.<sup>10</sup> Perhaps this was occasioned by the apparent lack of a native word in Abkhaz for 'cymbal'. However, it would be possible to remain closer to the original by, for

<sup>7</sup> NT 1963: 422.

<sup>8</sup> Dzots'enidze / Danelia 1974: 144.

<sup>9</sup> Interestingly, Classical Armenian *hamarēi* (Bible 1805: 423) could be the Imperfect either of what in the Present would be *hamarem* 'I reckon' or of what in the Present would be *hamarim* 'I am reckoned'. However, both the modern languages have active verbs, viz. Western Armenian *hamarum unēi* 'I had reasoning' (Bible 1955: 165) vs Eastern *ēi hamarum* 'I used to reason' (Bible 1974: 1307).

<sup>10</sup> Lasuria 2004: 395–396.

example, adopting the sequence offered Recommendation 1 in the Appendix,<sup>11</sup> which means ‘My voice resembles an echoing brass or the clanging that issues from dishes/disks being banged together’.

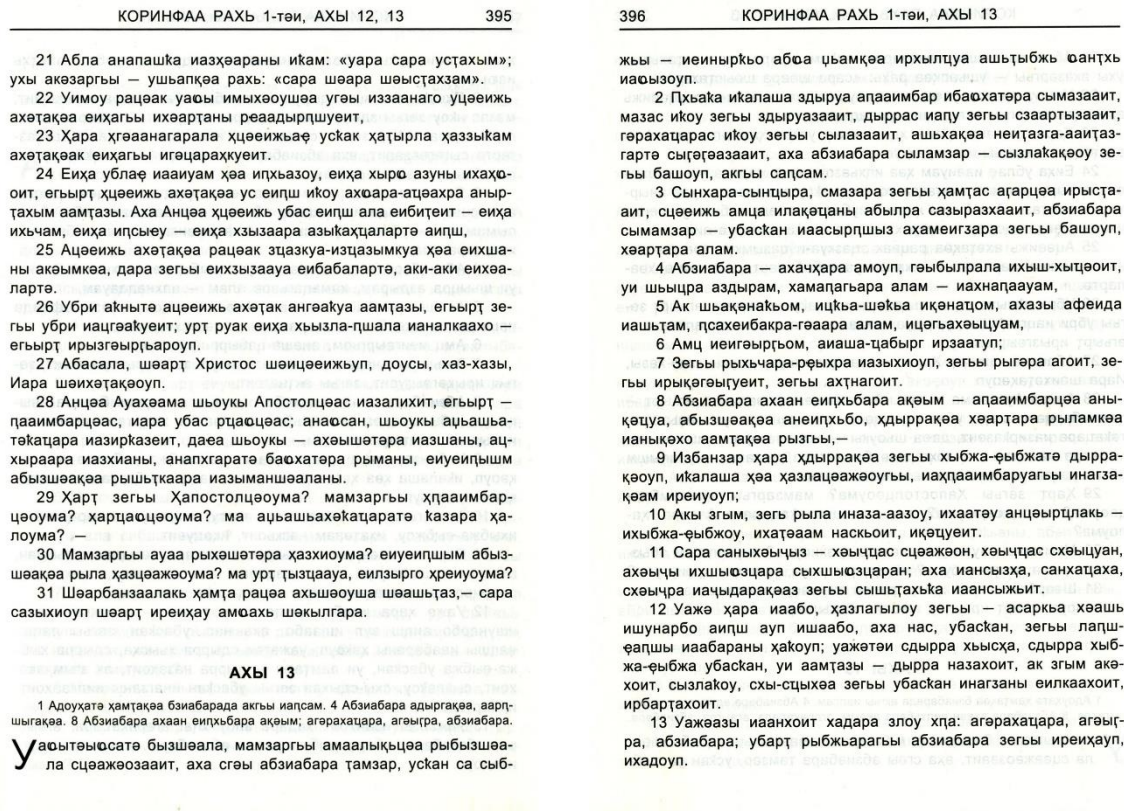


Fig. 3: Paul's First Epistle to the Corinthians ch. 13 in Lasuria's Abkhaz translation

Rather than reduction, it is expansion that we tend to find throughout this translation, as first exemplified in Verse 2 when the ability to move mountains is made dependent on the added characteristic of strength (*сәв<sup>w</sup>в<sup>w</sup>аза:jt* = *сыҕәҕәазааит* ‘may I be strong’), rather than the mere possession of faith, whilst the assertion of having no value is amplified by the inserted statement of capabilities being useless (viz. ‘All my capabilities are in vain’).

b) The first two verbs in Verse 5 (‘It does nothing unbecoming – it smashes nothing to pieces’) must be an attempt to make specific the Greek’s more general characterisation of Love’s ‘not behaving shamefully’, though the original’s non-specificity could be conveyed in the verb-phrase suggested in Recommendation 2 (‘It does not behave inappropriately’).

c) Moving to Verse 9, Lasuria’s rendering of the original’s uncomplicated use of just the two verbs ‘we know’ and ‘we prophesy’ with the qualification of each by ‘in part’ seems exceptionally wordy in its expansion.

d) Verse 10 is even more fulsome in its expansion – the abstraction ‘the perfect’ is represented by two relativised verbs, whilst its antonym ‘the imperfect’ is interpreted in three such verb-

<sup>11</sup> The recommendations were provided by my wife, Zaira Khiba.

forms. The verse then ends with two verbs ('move aside/make way' and 'disappear'), whereas the original makes do with just the one. This stylistic device of utilising two words to describe what is basically a single concept is known by the Greek term *hendiadys*, literally 'one through two', and it would seem to be a favourite of this particular author.<sup>12</sup>

e) In Verse 11 'when I became a man' is supplemented by the pre-posed additional clause 'when I grew up'.

f) Verse 12's current 'part-knowledge' is subjected to the now expected expansion, whilst the final four [*sic!*] verbs used by Lasuria all relate to the future, which means that the comparison 'I shall know/understand/perceive just as I was/am known/understood/perceived' is missing. Could this *lacuna* result from uncertainty about the reason for the original's temporal opposition, or did the source of the translation only refer to the future? Whatever the reason, the omission can easily be remedied by the three words in Recommendation 3, which straightforwardly state 'I shall understand (it/them) as I was understood'. Interestingly, the Georgian PV also has only the future (viz. 'I shall know/perceive as I shall be known/perceived'),<sup>13</sup> whereas the IBT versions agree on the stative Present form 'I am known/perceived' (*var shetsnobili* = ვარ შეცნობილი). One Russian Bible I consulted (2011) amplifies the text in its own seeming attempt at clarification by translating as *Тогда я буду знать так же полно, как знает меня Бог* 'Then I shall know just as completely as God knows me'!<sup>14</sup>

Unsurprisingly, the Abkhaz closes with yet another example of amplification combined with a *hendiadys* ('Love is the greatest of all, it is the main one').

### 3. Discussion

And so, we have before us different approaches to translating this passage from the New Testament. What do readers want from a translation? Do they want it to stay as close to the original as the receiving language permits, something that might be felt to be especially important for translators of biblical texts, or do they perhaps prefer something a little 'lighter' or perhaps 'more readable' with added clarification where necessary? I recall back in the 1980s when I presented one of my informants with a Georgian NT (something of a rarity in those late Soviet times), his reaction was that he found it 'somewhat boring'! As we say in English, each to his/her own.

---

<sup>12</sup> For a discussion of what he terms 'synonym pairs' see Boeder (1991).

<sup>13</sup> Old Georgian has: ვიცნა, ვითარცა შევემეცნე, with a variant-reading for the latter verb: ცნობილ ვიქმენ (Dzots'enidze / Danelia 1974: 144, footnote CD 12).

<sup>14</sup> The English rendition 'For now we see through a glass darkly' in the Authorised Version (*Bible* 1611) is one of those expressions which has so embedded itself in the language that it is widely known and quite often quoted. The Greek literally says: 'For yet we see through a looking-glass in an enigma/mirror/lens enigmatically', which does not sound quite right in English but could be slightly modified to a perfectly acceptable 'For yet/now we see through a looking-glass/mirror/lens enigmatically', retaining a form of the *very* word found in the Greek (ἀνύμω) which, of course, has long existed in English as the borrowing *enigma*. Even if this would suffice, it does not feel as poetically powerful as the words chosen by the 1611 translators. For comparison *The New English Bible* (1970) has 'Now we see only puzzling reflections in a mirror'.

I shall end by explaining why I chose this text for examination. When I was at secondary school throughout the 1960s, each morning began with a whole-school assembly, which included a short reading from the Bible in the Authorised Version of 1611. And the conclusion of this Epistle happened to stick in my head in these words: ‘And now abideth faith, hope, charity, these three, but the greatest of these is charity.’ At some point in later years I became conscious when hearing/reading the quote that ‘love’ was substituting for ‘charity’. Why?

As we see from Fig. 1, the Greek word for the human quality lauded by St. Paul is ἀγάπη, rather than either of the other two words the language possessed to express the different notions of ‘love’, namely ἔρωσ and φιλία. How are these words differentiated? The 8<sup>th</sup> edition of Liddell & Scott’s massive *Greek-English Lexicon* defines each as follows:<sup>15</sup>

ἀγάπη: love, especially brotherly love, charity... the love of God for man and of man for God

ἔρωσ: love, mostly of the sexual passion

φιλία: friendly love, affectionate regard, fondness, friendship.

In St. Jerome’s Latin Vulgate translation of the NT, ἀγάπη is mostly rendered by *caritas*, which ultimately gave English *charity*, but sometimes by *dilectio*. However, throughout our epistle it is *caritas* that is used. With regard to early renditions in English, the large Oxford English Dictionary says the following under its entry for ‘charity’:<sup>16</sup>

[T]he 16<sup>th</sup> cent[ury] Eng[lish] versions from Tindale to 1611, while rendering ἀγάπη sometimes “love”, sometimes “charity” did not follow the *dilectio* and *caritas* of the Vulgate, but used “love” more often (about 86 times), confining “charity” to 26 passages in the Pauline and certain of the Catholic Epistles... In the Revised Version 1881, “love” has been substituted in all these instances, so that it now stands as the uniform rendering of ἀγάπη.<sup>17</sup>

So, if the word ‘charity’ is essentially defined today as ‘Without any special Christian associations: Love, kindness, affection, natural affection: now esp[ecially] with some notion of generous or spontaneous goodness’, in the 16<sup>th</sup> century it was more ‘The Christian love of our fellow-men; Christian benignity of disposition expressing itself in Christ-like conduct’. Indeed, it would seem that William Tyndale himself even in the 1520s felt that ‘charity’ was more linked to the ‘giving of alms’, sensing the shift to the modern sense of charitable giving.

We, thus, see here why the use of ‘love’ by Tyndale in his translation of the NT (specifically in our text) was so controversial. We are able to read what he actually wrote thanks to the survival of only three copies of his 1526 English NT (see Fig. 3), so successfully repressed was the print-run because of the Crown’s aversion to the very existence of a vernacular NT (or Bible) in what was still a Catholic country at that time, Latin then being the *lingua franca* among educated classes across Europe. So furious was the wide-ranging and vicious doctrinal dispute between Tyndale and Sir Thomas More, who upheld the Catholic position (and paid with his life for it in 1535), that Tyndale had to take refuge on the continent, where he was finally run to ground and murdered in 1536. Manifestly, words have (sometimes fatal) consequences, and one needs

<sup>15</sup> Liddell / Scott 1901: 6 / 583 / 1672.

<sup>16</sup> OED 1971: 289–290.

<sup>17</sup> The references are to William Tyndale (c. 1494–1536), whose New Testament in English was published in 1526, the King James I’s Authorised Version of the Bible in English of 1611, and the Revised edition of the NT published in Oxford in 1881.



to be ever sensitive to how others are likely to interpret the nuances of what one says and/or writes. Ironically, a mere three years after Tyndale's death, King Henry VIII himself sanctioned the printing of his Great English Bible!

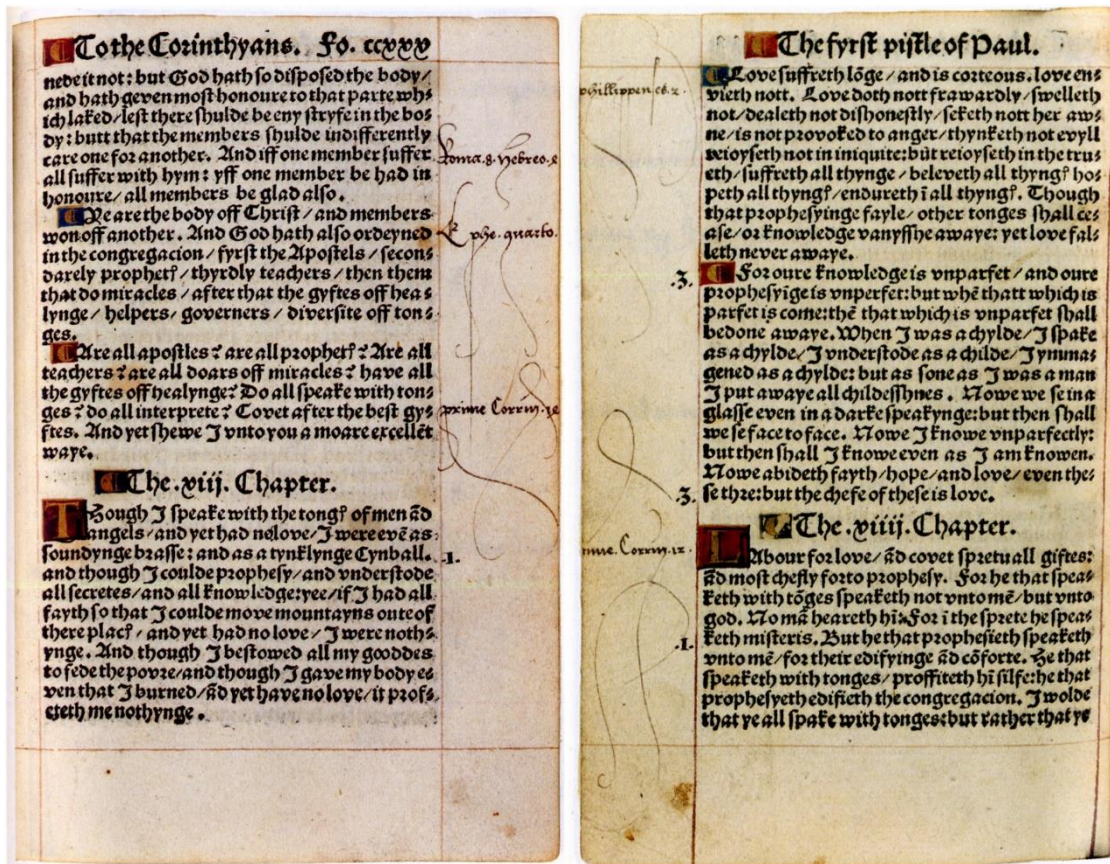


Fig. 4: Paul's First Epistle to the Corinthians ch. 13 in Tyndale's "The Newe Testamente", 1526, 230rv

#### 4. Conclusions

This, then, completes our consideration of the Georgian and Abkhaz translations of St. Paul's 1<sup>st</sup> Epistle to the Corinthians ch. 13, being a paean to ἀγάπη, with an added detour on the rendition of the Greek term in English – be it charity, be it love. And, given the state of the world, regardless of religious or ethnic differences, we can surely agree that the lesson that St. Paul was preaching in this text is perhaps even more relevant today than it was when originally delivered 2,000 years ago.

#### References

- Bible (1611): *The Holy Bible, Conteyning the Old Testament, and the New. Newly Translated...* Appointed to be read in Churches, London: Robert Barker. <https://archive.org/details/1611TheAuthorizedKingJamesBible/>.
- Bible (1805): *Աստուածաշունչ մատենան հին և նոր կտակարանաց...* Յաշխատասիրութենէ Տեանն Հ. Յովհաննու Չօրսպեան, հատոր 4, Վենետիկ: Սր. Ղազարու.

- Bible* (1955): *საბუთადგონი*. Vienna.
- Bible* (1970): *The New English Bible with Apocrypha*. Oxford / Cambridge: Oxford University Press / Cambridge University Press.
- Bible* (1974): *საბუთადგონი*. Tehran: Bible Society in Iran. Reproduced by photography from the 1896 edition.
- Bible* (1989): ბიბლია [*Bible*], თბილისი: საქართველოს საპატრიარქო [Patriarchate of Georgia].
- Bible* (1989–1990): ბიბლია. ძველი და ახალი აღთქმა. პირველი საწდელი გამოცემა [*Bible. Old and New Testaments. First Experimental Publication*], 4 vols, Stockholm: Institute for Bible Translation.
- Bible* (2002): ბიბლია. ძველი და ახალი აღთქმა [*Bible. Old and New Testaments*], Stockholm: Institute for Bible Translation.
- Bible* (2011): *Библия. Книги священного писания Ветхого и Нового Завета (Канонические)* [*Bible. The Books of the Holy Scripture of the Old and New Testament (Canonical)*], Moscow: Russian Bible Society.
- Boeder (1991): Winfried B., “Note on Synonymic Parallelism and Bilingualism”, *Studia Linguistica* 45 (1/2), 97–126.
- Dzots’nidze & Danelia (1974): პავლეს ეპისტოლეთა ქართული ვერსიები. გამოსაცემად მოამზადეს ქეთევან ძოწენიძემ და კორნელი დანელიამ. აკაკი შანიძის რედაქციით [Georgian Versions of the Pauline Epistles. Prepared for publication by Ketevan Dzots’nidze and K’orneli Danelia. Edited by Ak’ak’i Shanidze], თბილისი: მეცნიერება. <https://digitallibrary.tsu.ge/book/2021/july/shromebi/pavles-epistoleta-qartuli-versiebi.pdf>.
- Lasuria (2004): Мутьни Лашәриа, *Ауацаთჿ ჳხუ* [*The New Testament*]. 2004.
- Liddell & Scott (1901): Henry George L., Robert S., *Greek-English Lexicon*. Eighth edition, revised throughout. New York / Chicago / Cincinnati: American Book Company. <https://archive.org/details/greekenglishlex00lidduoft>.
- NT* (n.d.): *Bagster’s Critical New Testament. Greek and English*, London: Bagster.
- NT* (1526): *The Newe Testamente* [William Tyndale], Worms: Peter Schöffer. <http://digital.wlb-stuttgart.de/purl/bsz35126731X>. Facsimile edition: *The New Testament: A Facsimile of the 1526 Edition*. Translated by William Tyndale, with an Introduction by David Daniell, London: British Library.
- NT* (1881): *The Holy Bible. 5. The New Testament of our lord and saviour Jesus Christ...* being the version set forth A.D. 1611 compared with the most ancient authorities and revised. Oxford: Oxford University Press.
- NT* (1963): ახალი აღთქმა უფლისა ჩვენისა იესო ქრისტესი [*The New Testament of Our Lord Jesus Christ*], Tbilisi: Patriarchal Council.
- NT* (1980/1991): ახალი აღთქმა და ფსალმუნები. გადამუშავებული გამოცემა [*New Testament and Psalms. Re-worked Edition*], Stockholm: Institute for Bible Translation.
- NT* (1982): ახალი აღთქმა უფლისა ჩვენისა იესო ქრისტესი და ფსალმუნი. განახლებული გამოცემა [*The New Testament of Our Lord Jesus Christ and Psalms. Updated Publication*], Stockholm: Institute for Bible Translation.
- OED* (1971): *The Compact Edition of the Oxford English Dictionary*. Complete Text Reproduced Micrographically in two volumes, vol. 1, Oxford: Oxford University Press.

*Vulgata* (2007): *Biblia Sancta Vulgata*. Editio Quinta (Robert Weber, Roger Gryson), Stuttgart: Deutsche Bibelgesellschaft.

### Appendix 1: Abkhaz Text with Interlinear Translation

1. *Уафытәыџсатә* *бызшиә́ла* *мамзърҭы* *амаалықьҭәа* *рыбызшиә́ла*  
 human-ADJ language-INST or-even angel-PL their-language-INST  
*сцә́ажәо́зааит* *ах́а* *сгәы* *абзи́абара* *тә́амзар* *у́скан* *са*  
 may-I-be-speaking but my-heart love if-it-is-not-in-it then I/me  
*сыбжы́ы* *иени́ркьо* *абџа* *мьа́мқәа* *ирхы́лҭуа*  
 my-voice which-they-bang-together brass dish-PL which-emanates-from-them  
*а́шьты́бжъ* *аантхъ* *иаџы́зоуп*  
 the-sound dull it-resembles-it

#### [RECOMMENDATION 1

*ибжы́ныџуа* *абџа* *ма* *иени́ркьо* *ацьа́мқәа*  
 which-echoes brass or which-they-bang-together dish-PL  
*ирхы́лҭуа* *аи́нкъабжъ* *иаџы́зоуп*  
 which-emanates-from-them the-clanging it-resembles-it]  
 2. *ҭхьа́ка́* *и́кала́ша* *зды́руа* *аҭааимба́р* *ибаџа́тәра*  
 in-the-future which-is-to-happen who-knows-it the-prophet his-talent  
*сы́мазааит* *маза́с* *и́коу* *зегьы́* *зды́руазааит* *ды́ррас*  
 may-I-not-have-it as-a-secret which-is зегьы́ may-I-know-it as-knowledge  
*иаџу́* *зегьы́* *сзаарты́зааит* *гәраха́йара́с* *и́коу* *зегьы́*  
 which-is-about everything may-it-be-open-for-me as-faith which-is everything  
*сы́лазааит* *а́шьхақәа* *неи́тә́зга-ааи́тә́згартә*  
 may-it-be-in-me mountain-PL such-that-I-move-them-thither-hither  
*сы́ҭәҭә́азааит* *ах́а* *абзи́абара* *сы́ламзар* *сызла́кақәоу*  
 may-I-be-strong but love if-it-is-not-in-me things-of-which-I-am-capable  
*зегьы́* *ба́шоуп* *акгьы́* *саҭса́м*  
 all they-are-in-vain one-even I-am-not-worth-it

3. *Сынха́ра-сынйы́ра* *сма́зара* *зегьы́* *ха́мтҭас* *аба́рцәа*  
 my-property my-possession all as-gift the-poor-PL  
*иры́сҭааит* *сцәеи́жъ* *а́мца* *ила́қәйаны* *абы́лра́*  
 may-I-give-them-to-them my-body the-fire having-placed-it-on-it burning-of-it  
*сазы́разхааит* *абзи́абара* *сы́мамзар* *уба́скан*  
 may-I-become-at-ease-with-it love if-I-do-not-have-it then  
*иаа́сырҭы́шыз* *ахамеи́гзара* *зегьы́* *ба́шоуп* *ха́ртҭара́* *ала́м*  
 which-I-displayed not-sparing-oneself all it-is-in-vain advantage it-is-not-in-it  
 4. *Абзи́абара* *аха́чхара* *а́моуп* *гәыбы́лра́ла* *ихы́ш-хы́йҭәоит* *уи* *шьы́цра*  
 love patience it-has-it affection it-overflows it envy

*აზდირამ* *хамаңагыара алам* *иахнағаауам*  
it-does-not-know-it arrogance it-is-not-in-it it-does-not-overdo-it=act showily

5. *Ак* *шьақәнакьом* *иқкьа-шәкьа* *иқәнайом* *ахазы*  
one-thing it-does-not-do-it-awry to-smithereens it-does-not-ravage-it for-itself  
*აფეიდა* *иашьтам* *ңсахейбакра-гәаара алам* *ицәгьахәыцуам*  
profit it-does-not-seek-it raging-anger it-is-not-in-it it-does-not-think-evil

[RECOMMENDATION 2

*Ишакәым* *ахы* *мѡаѡнагѡм*  
how-it-is-not itself it-does-not-conduct-it = it does not behave inappropriately]

6. *Амц* *иειгәырѡом* *аи́аша-ѡабырг* *ирз́аатун*  
lie it-does-not-rejoice-in-it truth-honesty it-is-open-to-them

7. *Зегьы* *рыхьчара-рѣыхра* *иазыхи́оуп* *зегьы* *рыгарá*  
all their-protection-their-saving it-is-ready-for-it all their-heart(ness)

*агѡит* *зегьы* *иры́қәгәыѡуеит* *зегьы*  
it-carries-it [= it believes them all] all it-has-hope-in-them all/everything

*ахт́нагѡит*  
it-endures-it

8. *Абзи́абара* *аха́ан* *еиѡхьба́ра* *а́қәым* *аѡааимба́ррақәа*<sup>18</sup>  
love ever dissipation it-is-not-on-it prophecies

*аны́қәйуа* *абызи́ақәа́* *анейѡхьбо* *ҳды́ррақәа*  
when-they-vanish languages when-they-dissipate our-knowledge(s)

*хәарт́арá* *ры́ламкәа* *ианы́қәхо* *а́амт́ақәа* *рызгьы*  
advantage it-not-being-in-them when-they-remain the-times for-them-too

9. *Изба́нзар* *ҳарá* *ҳды́ррақәа* *зегьы* *хы́бжа-ѣы́бжатә*  
because we/us our-knowledge(s) all incomplete

*ды́ррақәоуп* *ика́лаша* *ҳәа* *ҳазлацә́жәо[у]гьы*<sup>19</sup>  
they-are-knowledge(s) what-will-happen saying and-what-we-speak-of

*иахѡааимба́руагьы* *инагзақә́ам* *ирейуоуп*  
and-what-we-prophecy thing-which-are-unfulfilled they-are-of-them

10. *Акы́* *згым* *зегь* *ры́ла* *иназа́-аазо́у*  
one-thing which-does-not-lack-it all in-respect-of-them what-is-developed

*иха́тәаау*<sup>20</sup> *анцә́ыры́лакь* *ихы́бжа-ѣы́бжоу* *иха́тәаам*  
what-is-brim-full whenever-it-emerges what-is-half-complete what-is-not-brim-full

*наскьѡит* *и́қәйуеит*  
it-moves-over it-disappears

<sup>18</sup> The published text has *аѡааимба́рца* ‘the prophets’, but we need ‘prophecies’, as the revised text now reads.

<sup>19</sup> The bracketed letter would be appropriate in a Stative verb, but, as this verb is Dynamic, it should not be present.

<sup>20</sup> The published text has *ихаатәу*, based on the root for ‘sweet’, which makes no sense. I am grateful to Gьында Kʷʷytʷnia [Gunda Kvitsinia] and Asida Alamia for suggesting this correction.



11. *Сарá саны́хэычыз хэычѳ́ас сѳ́а́жэон хэычѳ́ас схэ́ѳуан*  
 I when-I-was-a-child as-a-child I-used-to-speak as-a-child I-used-to-think  
*ахэычѳ́ы ихи́ѳѳ́а́ра сыхи́ѳѳ́а́ран а́хá иансы́зх́а санх́а́ѳаха*  
 child his-mind it-was-my-mind but when-I-grew when-I-became-a-man  
*схэ́ычр́а иа́чѳ́ыдарақ́аз зегь́ы сѳ́ишьт́ахь́ка иаансы́жьѳит*  
 my-childhood what-were-its-particularities all behind-me I-left-them
12. *Уажэ х́арá иа́або х́азлагь́лоузе́гы́ асáркья х́аашь*  
 now we what-we-see in-which-we-stand everything mirror cloudy  
*иш́унарб́о аиѳ́и ауп иша́або а́хá нас уба́скан*  
 how-it-shews-it/them-to-you like-it it-is how-we-see-it/them but then then  
*зегь́ы лаѳ́и-ѳаѳ́ишы́ иаабараны́ х́а́коуп уажэ́тáи сдѳ́рра*  
 everything face-to-face destined-that-we-see-it/them we-are of-now my-knowledge  
*хьысх́а́ сдѳ́рра хьыб́жа-ѳы́б́жаубáскан уи а́амт́азы ды́рра*  
 weak my-knowledge half-complete then that at-the-time knowledge  
*назахо́ит ак згь́ым а́кэхоит сы́злакóу*  
 it-becomes-perfect one-thing which-does-not-lack-it it-will-become my-capacity  
*схы-сѳ́ы́хэа зегь́ы уба́скан ина́заны́ (и)еилкаахо́ит*  
 my-head-my-tail all then perfectly it/they-will-be-understood  
*ирб́арт́ахоит*  
 it/they-will-become-manifest

[RECOMMENDATION 3

*сарá сие́илкааз (и)еилы́скаауеит*  
 I as-I-was-understood I-shall-understand-it/them]

13. *Уажэ́азы́ианхо́ит х́адарá злоу х́ѳа агэ́рахаѳ́арá агэ́ы́ѳра*  
 for-now they-remain primacy in-which-it-is three faith hope  
*абз́и́абара уба́ртѳ ры́б́жъя́рагы́ абз́и́абара зегь́ы иреи́х́а́уп*  
 love those and-among-them love all it-is-the-greatest-of-them  
*ихадо́уп*  
 it-is-the-main-one

**Appendix 2: Translation of ML's I Corinthians 13**

1. Should I be speaking in human language or the language of angels, but if love is not in my heart, then my voice resembles the dull sound that emanates from bronze plates clashing against each other.

2. Should I have the gift of a/the prophet who knows what is destined to happen in the future, were I to know all secrets, were everything there is to be known open to me, were everything there is by way of faith within me, were I strong enough to convey mountains hither and thither, but if love is not within me, all my capabilities are in vain, I have no value.

3. Were I to give as a gift to poor people all my property [and] my wealth, were I to place my body in the fire and become reconciled to the burning, if I do not have love, then all the selflessness I have displayed is in vain [and] there is no advantage in it.
4. Love has patience, it overflows with affection, it does not know envy, there is no arrogance in it, it does not behave overweeningly,
5. It does nothing unbecoming – it smashes nothing to pieces, it does not seek profit for itself, there is no rage-[and]-anger in it, it has no bad/evil thoughts,
6. It does not rejoice in a lie, it is open to truth-[and]-honesty;
7. It is ready to offer protection-[and]-salvation to all, it believes all, it has hope in all, it endures all things.
8. There is never any dissipation in love – even in times when prophecies vanish, when languages dissipate, when our areas of knowledge remain devoid of profit,
9. For all our areas of knowledge are half-complete areas of knowledge, and the things of which we speak as going to happen and what we prophesy belong to the realm of the unattained;
10. When that which lacks nothing, which is developed in all respects, which is brim-full emerges, that which is half-complete, that which is not brim-full will move aside [and] disappear.
11. When I was a child, I spoke as a child, I thought as a child, my mind was the mind of a child; but, when I grew up, when I became a man, I left behind me all things that were the particularities of my childhood.
12. Now, how we see everything that we see [in the world] in which we find ourselves resembles the way that a cloudy mirror shews it to you, but then, at that moment, we are destined to see everything face to face; my present weak knowledge, my halfcomplete knowledge will then become perfect knowledge at that time, it will become lacking in nothing, [and] all my capacity, my essential being shall then become perfectly understood [and] revealed.
13. For now there remain three primacies: faith, hope, love, and among those love is the greatest of all, it is the main one.

## პავლე მოციქულის პირველი ეპისტოლე კორინთელთა მიმართ

### თავი 13

ქართული და აფხაზური თარგმანების შედარება

(ბერძნული ორიგინალის გათვალისწინებით)

ჯორჯ ჰიუიტ(ი)

(SOAS, United Kingdom)

DOI: <https://doi.org/10.62235/dk.3.2024.8516>

[gh2@soas.ac.uk](mailto:gh2@soas.ac.uk) || ORCID: [0000-0002-7330-4107](https://orcid.org/0000-0002-7330-4107)

საყოველთაოდ არის აღიარებული, როგორც ბიბლია მსოფლიოში ყველაზე ხშირად თარგმნილი წიგნია. ქართულ ლიტერატურას დამწერლობის მრავალსაუკუნოვანი ტრადიცია გააჩნია, მათ შორის თარგმნილი ლიტერატურის. მათ შორისაა, რასაკვირველია, ბიბლიის (ძველი და ახალი აღთქმების წიგნებს) არაერთი სრული თარგმანი. ქართული ენისაგან განსხვავებით, აფხაზურ ენაზე ბიბლიის მხოლოდ ახალი აღთქმის ერთი თარგმანი არსებობს (სრული ვერსია), რომელიც რუსულიდან არის გადმოთარგმნილი და შესრულებულია მწერალ მუშნი ლასურიას მიერ. რამდენადაც ჩვენთვის ცნობილია, მუშნი ლასურიას თარგმანი კერძო ინტერესის საფუძველზე შესრულდა და გამოიცა 2004 წელს.

ბიბლიის თარგმანის ინსტიტუტი, რომლის დირექციაც ოდესღაც სტოკჰოლმში იყო დაბინავებული, ხოლო ახლა მოსკოვშია, რამდენიმე წლის წინ დამიკავშირდა და მთხოვა ლასურიას მიერ შესრულებული ოთხთავის აფხაზური თარგმანის რედაქტირება. აფხაზურ თარგმანზე მუშაობა დიდხანს გაგრძელდა და მრავალი თვის განმავლობაში ინტენსიური მუშაობა დამჭირდა იმისათვის, რომ მეჩვენებინა, თუ რა ცვლილებები იყო შესატანი თარგმანში, რომ აფხაზური თარგმანი ბერძნულ დედანთან დაგვეახლოვებინა. იმავდროულად ოთხთავის სხვა აფხაზურ თარგმანზეც ვმუშაობდი, კერძოდ, ზაირა ხიბას მიერ რუსულიდან აფხაზურად თარგმნილ ოთხთავზე, რომელიც იმავე პრინციპით გადავამუშავე (ზაირა ხიბას მიერ აფხაზურად თარგმნილი ოთხთავების ეს ვერსია ინტერნეტში ხელმისაწვდომია<sup>21</sup>).

აქედან გამომდინარე, გადავწყვიტე წინამდებარე სტატიის დაწერა, რომელშიც ახალი აღთქმის ასევე ძალიან ცნობილ ნაწილებში, კერძოდ, პავლე მოციქულის ეპისტოლეში კორინთელთა მიმართ, სიყვარულის თემას მინდა შევეხო. ამასთან, დისკურსის განხილვას ვაპირებ არა მარტო იმ თვალსაზრისით, თუ როგორ გადათარგმნეს აღნიშნული პასაჟი ქართველმა მთარგმნელებმა ქართულად (როგორც ძველ ქართულში, ისე ახალში), ხოლო მუშნი ლასურიამ აფხაზურად, არამედ იმ თვალსაზრისითაც, თუ რამდენად ესადაგება შესრულებული თარგმანები ბერძნულ წყაროს. საკითხის განსახილველად და

<sup>21</sup> <http://georgehewitt.net/articles/miscellaneous/316-the-four-gospels-in-abkhaz-translated-from-russian-by-zaira-kiazimovna-khiba>.

თარგმანების შესადარებლად გამოვიყენე ახალი აღთქმის ის გამოცემები და სამეცნიერო პუბლიკაციები, რომლებიც ჩამოთვლილია გამოყენებული ნაშრომების სიაში.

ზოგადად, პავლე მოციქულის პირველი ეპისტოლე კორინთელთა მიმართ, კერძოდ, მე-13 თავი, საყოველთაოდ განიხილება, როგორც სიყვარულის თემისადმი მიძღვნილი წერილი. შეიძლება გასაკვირი იყოს, მაგრამ ინგლისურ თარგმანებში ყველგან არ იკითხება სიტყვა 'love' (სიყვარული); მის ნაცვლად გამოყენებულია სიტყვა 'charity' (ქველმოქმედება).

სტატიაში ჩვენ განვიხილავთ, ერთი მხრივ, იმ მსურველ პოლემიკის მიზეზებს, რომელსაც ადგილი ჰქონდა მე-16 საუკუნის ინგლისში ამ ჩანაცვლების გამო, მეორე მხრივ კი, კრიტიკულად განვიხილავთ პავლე მოციქულის პირველი ეპისტოლეს (კორინთელთა მიმართ) აფხაზურ თარგმანს და წარმოვადგენთ ჩვენს რეკომენდაციებს.

დასასრულს, თუ გავითვალისწინებთ მსოფლიოში არსებულ მდგომარეობას, განურჩევლად რელიგიური თუ ეთნიკური განსხვავებებისა, ნამდვილად შეგვიძლია დავეთანხმოთ იმ აზრს, რომ გაკვეთილი, რომელსაც წმინდა პავლე ქადაგებდა ზემოთ განხილულ ტექსტში, შესაძლოა დღეს უფრო მეტად იყოს აქტუალური, ვიდრე თავდაპირველად, 2000 წლის წინ იყო.

# The Theoretical Grounds and the Practical Value of Distinguishing the Markers of Evidentiality in the Verb Paradigms of the Kartvelian Languages

Maia Lomia, Ketevan Margiani (Tbilisi)

(Ivane Javakhishvili Tbilisi State University)

DOI: <https://doi.org/10.62235/dk.3.2024.8517>

[maia.lomia@tsu.ge](mailto:maia.lomia@tsu.ge) || ORCID: 0009-0009-4695-7426

[ketevan.margiani@tsu.ge](mailto:ketevan.margiani@tsu.ge) || ORCID: 0009-0005-8571-3943

**Abstract:** Evidential verb forms of the Perfect are a common Kartvelian phenomenon. Expression of evidentiality in this case is a secondary function on the synchronic plane. Perfect evidential verb forms do not have a special morphological marker. They use lexical means to express that something has not been seen. In this way, they contrast with aorist verb forms expressing an action that was seen. Evidential verb forms of the Imperfect appear only in the non-literary Kartvelian languages (Megrelian, Laz and Svan). Out of these, in Laz, evidentiality is expressed lexically (in a descriptive way), whereas in Megrelian and Svan, evidentiality has a special morphological marker. In Megrelian, the Present and Imperfect verb forms express a seen action. They are opposed to evidential verbs in the Imperfect, which denote an unseen action. In Svan, the neutral Imperfect contrasts two forms: the Evidential Imperfect I in the superessive version and the Evidential Imperfect II. Originally, the Megrelian Evidential Imperfect I and Evidential Imperfect II, and the Svan Evidential Imperfect II were descriptive. As a result of transformation, these forms turned into organic formations and developed special markers of evidentiality. The Svan Evidential Imperfect I in the superessive version seems to have been organic from the very start. In Svan and Megrelian, evidentiality has developed as a morphological category. The segmentation and glossing of evidential verb-forms in the non-literary Kartvelian languages and the distinction of the markers of evidentiality serve important practical aims that are thematised in the present article.

**Keywords:** Evidentiality, Kartvelian Languages, morphological markers of Evidentiality

## Introduction

Evidentiality is a universal category. It is found in every language, although the linguistic means of its expression may vary even within related languages. In some languages, evidentiality is a grammatical category, whereas in others it is a lexical-semantic feature. In cases where evidentiality is a grammatical category, the chief means of its expression is the verb.<sup>1</sup> Evidential verb-forms can express both non-modal and modal semantics and are of two types: perfective and imperfective.

The Kartvelian languages yield especially interesting information on the comparison of literary and non-literary languages. The group of Kartvelian (South Caucasian) languages embraces Georgian, Megrelian, Laz and Svan. Out of these, Georgian is the literary language whereas Megrelian, Svan and Laz are non-literary.

The present paper focuses on the theoretical grounds of the formation and segmentation of non-modal evidential **perfective** and **imperfective** verb forms in the Kartvelian languages.

---

<sup>1</sup> For details, see Nitsolova 2007: 122.

## 1. Perfective evidential verb forms in the Kartvelian languages

Perfective evidential verb forms (represented by inversive resultatives) are a common Kartvelian phenomenon. In these forms, the expression of evidentiality is a secondary function, activated on the synchronic level. According to shared opinion, they have emerged from bipersonal stative verbs as a result of inversion. Their initial meaning was resultative, and they developed the semantics of **unseen** action later.

Perfective evidential verb forms are mostly devoid of a morphological marker, they express unseen action only lexically. In this regard, they are opposed to aorist verb forms expressing **seen** actions. In this respect, the situation is similar in three Kartvelian languages (Georgian, Megrelian and Laz) while Svan behaves differently in that here, transitive perfective verbs without a preverb have developed a special marker for unseen action: the suffix *-en*. In contrast, in forms with a preverb, the expression of resultativeness is overshadowed as these forms have acquired only the capacity of denoting unseen actions. Naturally, they are also opposed to verb forms of the aorist group. Cf. the following examples:

Georgian:

(1) *da-çer-a* (PRV-write-AOR.S3.SG) – *da-Ø-u-çer-i-a* (PRV-O3-OV-write.EV\_PRF.S3.SG)

Megrelian:

(2) *do-çar-u* (PRV-write-AOR.S3.SG) – *du-Ø-u-çar-u-(n)* (PRV-O3-OV-write.EV\_PRF-THM-(S3.SG))

Laz:

(3) *do-ça(r)-u* (PRV-write-AOR.S3.SG) – *Ø-u-ça(r)-u-n* (O1-OV-write.EV\_PRF-THM-S3.SG)

Svan:

(4) *ad-iyr-e* (PRV-write-AOR.S3.SG) – *x-a-yr-en-a* (O3-SUPV-write-EV\_PRF-S3.SG)

(5) *ad-iyr-e* (PRV-write-AOR.S3.SG) – *ot-iyr-a* (PRV+O3-write.EV\_PRF-S3.SG)

Thus, all the Kartvelian languages are similar from the viewpoint of the genesis of their Perfect tenses: perfective verbs are derived from stative verbs, and their primary function is to denote the result of an action. As for their secondary function, the category of evidentiality, the Svan language reveals a different situation.

## 2. The issue of naming perfective evidential verb forms in Svan

The Perfect tenses of Svan, also termed ‘screeves’,<sup>2</sup> are given different names in the scholarly literature. This diversity is due to the fact that different scholars use different criteria. In some cases, the names of tenses differ even in the works of one and the same author. In order to denote the Perfect tenses, Varlam Topuria uses the terms ‘Past Perfect’ and ‘Previous Past’,<sup>3</sup> although, in some works, he names these tenses as ‘Resultative I’ and ‘Resultative II’.<sup>4</sup> Aleksandre Oniani and Zurab Chumburidze refer to the same tenses as ‘Evidential I’ and

<sup>2</sup> ‘Screeve’ (mçkrivi) is a term introduced by Akaki Shanidze (1973/1980: 215). It denotes a verb paradigm that is characterised by the grammatical categories of tense and mood. The English term was coined by Howard Aronson (1982: 41).

<sup>3</sup> ‘ნამყო სრული’ and ‘ნამყო წინარეწარსული’ (Topuria 1967a: 169 /173).

<sup>4</sup> ‘Результативное I’ and ‘Результативное II’ (Topuria 1967b: 85 / 2002: 78; 1985: 2002: 121).

‘Evidential II’.<sup>5</sup> Based on systematic research on the category of evidentiality and a comparison of the Kartvelian data, the tenses under analysis are here called ‘Evidential Perfect I’ and ‘Evidential Perfect II’.<sup>6</sup>

### 3. Imperfective evidential verb forms in the non-literary Kartvelian languages

Imperfective evidential verb forms have only emerged in the non-literary Kartvelian languages (Megrelian, Laz and Svan). Out of these, evidentiality is expressed lexically (in a descriptive way) in Laz, whereas in Megrelian and Svan, it has a special morphological marker. Thus, the issue of an identification of markers concerns only Megrelian and Svan.

#### 3.1 Imperfective evidential verb forms in the Megrelian language

In Megrelian, verb forms of the Present and Imperfect express a seen action. They are opposed to Evidential Imperfect forms, which denote an unseen action as in the following examples:

(6) *čar-un-s* (write-THM-PRS.S3.SG) – *no-čar+u-e-(n)* (EV\_PRF\_1-write-EV\_PRF\_2-S3.SG)

(7) *čar-un-d-u* (write-THM-EM-IMPF.S3.SG) – *no-čar+u-e-d-u* (EV\_PRF\_1-write-EV\_PRF\_2-EM-S3.SG)

The Megrelian imperfective evidential forms (‘Evidential Imperfect I’ and ‘Evidential Imperfect II’) were analytical at the beginning. As a result of transformation, these forms turned into an organic formation and developed special markers of evidentiality.

##### 3.1.1 The issue of naming the imperfective evidential verb forms in Megrelian

In the Georgian scholarly literature, the evidential forms are named differently because, when selecting the terms, priority was attached to either the form or the function. Giorgi Rogava comprised the imperfective evidential verb forms (paradigms with the prefix *no-*) under the common designation *nakvti* ‘feature, form’;<sup>7</sup> he considered that these forms should be regarded as a separate series IV, whereas the individual screeves should be called ‘Evidential III’ and ‘Evidential IV’,<sup>8</sup> because in the grammars of Kartvelian languages, the screeves of the III series denoting unseen actions were already termed ‘Evidential I’ and ‘Evidential II’. According to Izabela Kobalava, the imperfective evidential verb forms should be included in series I as they are based on the present tense and are opposed to the other screeves of the series as referring to unseen actions. Based on their content, Kobalava styled the imperfective screeves ‘Present Evidential’ and ‘Past Imperfect Evidential’.<sup>9</sup> Recently, the above-mentioned screeves were named ‘Evidential Imperfect I’ and ‘Evidential Imperfect II’.<sup>10</sup>

##### 3.1.2 The dynamics of the development of organic forms in Megrelian

In Megrelian, the imperfective evidential verb forms with the prefix *no-* are new formations as compared to other screeves. Initially, their structure was analytic in that they were based on a

<sup>5</sup> ‘I თურმეობითი’ and ‘II თურმეობითი’ (Oniani 1998: 202–203; Chumburidze et al. 2007: 136).

<sup>6</sup> For details, see Margiani et al. 2019: 74–76.

<sup>7</sup> The terms *nakvti*, also used by Arnold Chikobava (1962), and *mčkrivi* can be regarded as being synonymous.

<sup>8</sup> Rogava 1953: 17–18.

<sup>9</sup> ‘აწმეოს თურმეობითი’ and ‘ნამეო უწვევებლის თურმეობითი’ (Kobalava 2001: 132).

<sup>10</sup> Margiani et al. 2019: 74–76.

participle with the circumfix *no- -u* (e.g., \**no-čkir-u* ‘cut’) in combination with forms of the auxiliary verb ‘to be’ in the Simple Present or the Past Continuous, *ore(n)* ‘is’ and *ordu* ‘was’, as in \**no-čkir-u + ore(n)* ‘is (being) cut’, \**no-čkir-u + ore-d-u* ‘was (being) cut’.<sup>11</sup>

The final *-u* in the circumfix *no- -u* is a characteristic feature of the masdar in Zan (i.e., Megrelian and Laz); cf. *o-xačk-u* ‘to hoe’, *o-čar-u* ‘to write’ and so on.<sup>12</sup> In modern Megrelian speech, we find substantivised participles with a prefix *no-* such as *no-čkeri* ‘cutting’; besides, *no-* forms past participles as in *no-xačk-u-er-i* ‘fee for having hoed’.<sup>13</sup>

As a result of the structural-semantic transformation of imperfective evidential verb forms, we have, e.g.:

- (8) \**no-čkir-u + ore(n)* ‘(sth.) is cut’ > *no-čkir-u-e*<sup>14</sup> ‘(sbd.) is evidently cutting (sth.)’ (Evidential Imperfect I);
- (9) \**no-čkir+u + \*oredu*<sup>15</sup> ‘(sth.) was cut’ > *no-čkir+u-e-d-u* ‘(sbd.) was evidently cutting (sth.)’ (Evidential Imperfect II).

In the Megrelian language, there are subjective (with the marker /v/) and objective (with the marker /m/) paradigms of conjugation of the above-mentioned verbs. It should be noted that one of the features of evidentiality is the generalisation of person. Thus, the most productive forms in the language are the third person forms of such verbs. Formally, they are alike and are not distinguished by subjective and objective markers.

The transformation has affected both form and content. On the synchronic level, the transformed forms are imperfective evidential forms of organic (synthetic) formation, containing a circumfix *no- -e*, which is a **marker of evidentiality**.<sup>16</sup> There are several arguments in favour of this opinion:

- a) The prefix *no-* expresses the semantics of the past: “It seems that the shift of the time of the action into the past is supported by the prefix *no-* (< *na-*), used in nouns denoting former objects and in one of the forms of the past participle”.<sup>17</sup>
- b) The semantics of the evidential verb forms is defined by the past tense. The process of development of the semantics of evidentiality from Perfect tenses is universal, hence, it has developed almost equally in all Kartvelian languages.<sup>18</sup>
- c) The reduction of the Megrelian auxiliary verb *ore(n)* ‘is’ > *-e*: in Megrelian *-e*, together with *no-*, forms a circumfix expressing evidentiality, whereas the 3<sup>rd</sup> person marker in Megrelian is the suffix *-n*.

<sup>11</sup> Rogava 1953: 23.

<sup>12</sup> Chikobava 1936: 177–178 / 2008: 183–184.

<sup>13</sup> Rogava 1953: 23.

<sup>14</sup> In Megrelian, the auxiliary verb may be reduced from left to right and vice versa.

<sup>15</sup> With the sequence \**ore-d-u* > *or-d-u* > *-d-u*; *-d-* is a marker of continuous tense, and *-u* is the marker of a 3<sup>rd</sup> person subject.

<sup>16</sup> Kurdadze et al. 2017: 18.

<sup>17</sup> „მოქმედების დროის უკან გადატანას, როგორც ჩანს, ხელს უწყობდა ნო- (< ნა-) პრეფიქსი, რომელიც წინა ვითარების სახელებისაგან დროის მიმდევრობის ერთ-ერთ წარმოებაშია გამოყენებული“ (Kobalava 2001: 134 n. 1).

<sup>18</sup> Cf. Chikobava 1962; Arabuli 1984; Beridze 2009; Topadze 2011; Kurdadze et al 2017.



- d) Semantic and formal differences: imperfective evidential verb forms differ from forms of the simple and continuous Present in the semantics of unseen vs. seen action. Formally, the difference is marked by the circumfix *no-* – *-e*, and this very circumfix serves as a marker of evidentiality; cf.:

(10) *čaruns* ‘is writing’ (Simple Present) vs. *nočarue* ‘is evidently writing’ (*\*načerava*).<sup>19</sup>

(11) *čarundu* ‘was writing’ (Past Continuous) vs. *nočaruedu* ‘was evidently writing’.

- e) The suffix *-u* does not fulfil the function of a marker of evidentiality: the semantics of evidentiality (i.e., unseen action) was added to the organic (synthetic) form by the past tense prefix *no-*, the suffix *-u* did not express a tense meaning. On the synchronic level, it has survived in imperfective verbs, albeit devoid of its function.

### 3.2 Imperfective evidential verb forms in Svan

In underlining the complex nature of the Svan verb, Varlam Topuria noted the formation of screeves in Svan that differ from Georgian formations.<sup>20</sup> In the Svan verb conjugation, the Resultative is an especially complex screeve, due to its complicated formation and diversified semantics.

#### 3.2.1 The issue of naming the imperfective evidential verb forms in Svan

In Topuria’s treatise, the screeves peculiar of the Svan language belong to series I and are named ‘Evidential I’ and ‘Evidential II’. According to Aleksandre Oniani, these screeves should represent a separate series IV.<sup>21</sup> Later, the evidential verb forms of the first series – ‘Evidential I’ and ‘Evidential II’ – were termed ‘Evidential Imperfect I’ and ‘Evidential Imperfect II’.<sup>22</sup>

#### 3.2.2 Structural analysis of the imperfective evidential verb forms in Svan

In Svan, the neutral Imperfect is represented by two different formations: the ‘Evidential Imperfect I’ in superessive version and the ‘Evidential Imperfect II’. The form expressing the superessive version is characterised by the suffixes *-ina/-una*.<sup>23</sup>

(12) *xäbm-ina / -una* ‘sbd. was evidently binding (sth.) to sth.’

Varlam Topuria divided the suffixes *-ina / -una* into two parts as *-in-a / -un-a* but he said nothing specifically about the origin and function of their elements.<sup>24</sup> Oniani regarded *-ina/-una* as unseparable units.<sup>25</sup> According to later literature, the suffix *-un / -in* is a marker of evidentiality.<sup>26</sup> This opinion is based on several factors:

<sup>19</sup> A western Georgian dialectal form (Imeretian, Rachian).

<sup>20</sup> Topuria 1967: 18.

<sup>21</sup> Oniani 1998: 205. Cf. G. Rogava’s opinion regarding evidential verb forms in Megrelian (see 4.1.1 with n. 8 above).

<sup>22</sup> Margiani et al 2019: 74–76.

<sup>23</sup> Topuria 1967: 130–131.

<sup>24</sup> Topuria 1967: 130–131.

<sup>25</sup> Oniani 1998: 205–206.

<sup>26</sup> Margiani-Subari 2008–2009: 132.

- a) The Evidential Imperfect I is opposed to the Past Continuous; the semantic and formal markers of this opposition are reference to unseen action and the suffix *-un/-in*; cf.:
- (13) *xäbm-in-a /-un-a* ‘sbd. was evidently binding (sth.) to sth.’ vs. *xabämda* ‘was binding’.
- b) The suffix is also found in the Evidential Imperfect II. The meaning of the Evidential Imperfect II is precisely expressed by the forms of the type *naq̄vareba* ‘sbd. loved (sbd.)’ found in western Georgian dialects (Imeretian, Rachian); cf. the Megrelian Evidential Imperfect I with the prefix *no-*.

The Svan Evidential Imperfect II is especially complex from the structural viewpoint: it builds upon the Present stem, combined with a prefix *lə-ma-* and a suffix *-un-e*; in addition, the forms may appear with an additional copular verb. In the formation of this screeve, *-ma-* is often reduced to *-m-* due to the regular syncope of vowels in second syllables; as usual, the vowel is retained in the Lower Svan dialect of Lentekhi.<sup>27</sup> In the suffix *-un-e* the *-e* is often omitted; therefore, the preceding vowel appears either contracted or in an unlauded form. As a result of the phonetic change, the final form of the formant is then *-in* (*-un-e* > *-iin* > *-win* > *-in*). The forms without the final *-e* in the 3<sup>rd</sup> person are chiefly found without the auxiliary verb:

- (14) *\*lə-ma-dagr-un-e+li* vs. *\*lə-ma-dagr-un-e* > *lə-m-dagr-iin* > *lə-m-dag(w)r-in* ‘sbd. was killing (sbd.)’ (cf. Georgian *\*naklava*).<sup>28</sup>

Varlam Topuria considers *lə-* to be the prefix of a past participle<sup>29</sup> and *-ma-* to be a marker of active voice (Topuria 1967: 134–135). As for the prefix *m(a)-*, the scholar solely notes that “its function (like the function of the suffix *-un[/-in]*) is yet to be studied”<sup>30</sup> (Topuria 1967: 135). However, both V. Topuria and A. Oniani mention the formants *mə- – -e* and *mə-* as markers of the past participle that freely substitute the prefix *lə-* (Topuria 1967: 221; Oniani 1998: 269). In our opinion, in the screeve under analysis, *-mə-* is also obtained from a participle. The function of this marker in this case is to denote activity/duration of an action regardless of the voice of the verb.

In Georgian, Svan forms of the ‘Evidential Imperfect II’ are rendered analytically by using the Imperfect with the adverb *turme* (‘evidently, apparently’). It is well-known that the Georgian Imperfect expresses a continuous action; thus, this screeve denotes an ongoing process in the past.

The suffix *-une* is divided by Topuria into the elements *-un* and *-e*. As in other screeves, *-un-* here is a marker of evidentiality. Givi Matchavariani notes that the suffix *-un* is archaic,<sup>31</sup> thus matching our own opinion regarding the origin of this suffix; the differentiation of its function based on screeves must have emerged later. On the synchronic level, the semantic function of the suffix *-un-* is to express **evidentiality** (in the sense of unseen or improbable action).<sup>32</sup> Maybe

<sup>27</sup> E.g., in *ləmageli* ‘sbd. was evidently building’ (Topuria 1967: 131).

<sup>28</sup> Dialectal form (Imeretian, Rachian).

<sup>29</sup> Topuria 1967: 134.

<sup>30</sup> „-ე რომ -უნ- ელემენტის კუთვნილება არ არის, იქიდან ჩანს, რომ -უნ-ის გარეშეც შეიძლება ზმნაში გვეპოვნდეს: ბზ. ღებურდ-ე-ლი... ეცრ. ღებურდ-ე-ლი...; ... სხვა ფორმანტთა (-მა-, -უნ-ის) ფუნქცია სარკვევია“ (Topuria 1967: 134–135).

<sup>31</sup> Matchavariani 1974:131.

<sup>32</sup> Improbability is expressed in modal (epistemic-evidential) screeves which are not being discussed at this point.

at an earlier stage the suffix *-une* was either a stem ending of certain verbs or a functionally non-differentiated suffix.

#### 4. Conclusions

A systemic analysis of the verb paradigms has proven that among the Kartvelian languages, morphological markers of evidentiality are found only in Svan and Megrelian. Among these, the process of grammaticalisation of evidentiality is more vivid in Svan.

In Svan, there are special suffixes of evidentiality for verbs forms in the Perfect (*-en*) and Imperfect (*-un*). In Megrelian, however, a comparable suffix is found only in forms of the Imperfect. Hence we may conclude that

- a) in Svan, the ‘Evidential Perfect I’ and the ‘Evidential Imperfect I’ in superessive version were organic (synthetic) formations from the very start, whereas
- b) the Megrelian ‘Evidential Imperfect I’ and ‘Evidential Imperfect II’ as well as the Svan ‘Evidential Imperfect II’ were initially analytic and developed to be organic (synthetic) formations as a result of transformation. This was the time when a special marker of evidentiality appeared in both languages.

Thus, evidentiality has developed as a morphological category in Svan and Megrelian. An appropriate segmentation and glossing of evidential verb forms in the Kartvelian languages and the identification of the markers of evidentiality has theoretical and practical significance with a view to

- a) revealing the rich capacity of the language with regard to the expression of a universal category,
- b) describing the genesis of the category of evidentiality and proving its authenticity, and
- c) serving the purpose of a consistent morphological annotation of Megrelian and Svan texts that is usable for digital analyses.

#### References

- Arabuli (1984): ავთანდილ არაბული, მესამე სერიის ნაკვეთულთა წარმოება დამნიშვნელობა ძველ ქართულში. თბილისი: მეცნიერება. <https://iverieli.nplg.gov.ge/handle/1234/369145>.
- Aronson (1982): Howard I. A., *Georgian. A Reading Grammar*. Columbus, Ohio: Slavica. – Corrected edition, 1990.
- Beridze (2009): მარინე ბერიძე, „პერფექტულობის გამოხატვის ზოგი საკითხისათვის ქართულში“, in ალექსანდრე კარტოზია (ed.), გურამ კარტოზია / Guram Kartoziia 75 (ქართველოლოგიური ბიბლიოთეკა / Kartvelological Library, 11). თბილისი: სეზანი, 93–104.
- Chikobava (1936): არნოლდ ჩიქობავა, ჭანურის გრამატიკული ანალიზი ტექსტებითურთ / *Analyse grammaticale du dialecte tchane (laze) avec les textes*. ტფილისი: სსრკ მეცნიერებათა აკადემიის საქართველოს ფილიალის გამოცემა. Repr. in 'შრომები III, თბილისი: არნ. ჩიქობავას სახელობის ენათმეცნიერების ინსტიტუტი, 2008. <https://iverieli.nplg.gov.ge/handle/1234/10700>.

- Chikobava (1962): არნოლდ ჩიქობავა, „ქართული ზმნის ნაკვეთელთა დაჯგუფების პრინციპისათვის“, იბერიულ-კავკასიური ენათმეცნიერება 13, 93–107.
- Chumburidze et al. (2007): ზურაბ ჭუმბურიძე, ლელა ნიუარაძე & რამაზ ქურდაძე, სვანური ენა (გრამატიკული მიმოხილვა, ტექსტები, ლექსიკონი). ზურაბ ჭუმბურიძის რედაქციით. თბილისი: პეტიტი.
- Kobalava (2001): იზაბელა ქობალავა, „კავშირებითი კილო მეგრულში“, ენათმეცნიერების საკითხები 2001/4 (12), 111–143.
- Kurdadze et al. (2017): Ramaz K., Ketevan Margiani, Maia Lomia, “Category of Evidentiality in the Kartvelian languages: Problems, Research History, Perspectives”, *International Journal of Multilingual Education* 9, 12–22 <https://multilinguaeducation.openjournals.ge/index.php/ijml/article/view/6628/6621>.
- Matchavariani (1974): გივი მაჭავარიანი, „ასპექტის კატეგორია ქართველურ ენებში“, ქართველურ ენათა სტრუქტურის საკითხები 4, 118–143.
- Margiani et al. (2019): ქეთევან მარგიანი, რამაზ ქურდაძე, მაია ლომია, ევიდენციულობის კატეგორია ქართველურ ენებში / Ketevan M., Ramaz Kurdadze, Maia Lomia, *The Category of Evidentiality in the Kartvelian languages*. თბილისი / Tbilisi: ივანე ჯავახიშვილის სახელმწიფო უნივერსიტეტი. <https://rustaveli.org.ge/res/docs/b34eb2222c043ee03ff4910134182a7c2aeb47be.pdf>.
- Margiani-Subari (2008–2009), ქეთევან მარგიანი-სუბარი, ზემოსვანურ დიალექტთა მორფო-სინტაქსური ანალიზის ზოგი ასპექტი. თბილისი: მწიგნობარი. [Abstract: <http://eprints.iliauni.edu.ge/7565/>].
- Nitsolova (2007). Руселина Ницолова, “Модализованная эвиденциальная система болгарского языка”, in Виктор Самуилович Храковский (ред.), *Эвиденциальность в языках Европы и Азии. Сборник статей памяти Наталии Андреевны Козинцевой*. Санкт-Петербург: Наука, 107–196. <https://nenadict.iling.spb.ru/publications/1736>.
- Oniani (1998): ალექსანდრე ონიანი, სვანური ენა (ფონოლოგიისა და მორფოლოგიის საკითხები). (ქართველოლოგიური ბიბლიოთეკა, 2). თბილისი: სულხან-საბა ორბელიანის სახელობის თბილისის სახელმწიფო უნივერსიტეტის გამომცემლობა.
- Rogava (1953): გიორგი როგავა, „დრო-კილოთა მეოთხე ჯგუფის ნაკვეთები ქართველურ ენებში“, იბერიულ-კავკასიური ენათმეცნიერება 5, 17–31.
- Shanidze (1973): აკაკი შანიძე, ქართული ენის გრამატიკის საფუძვლები. I. მორფოლოგია. მეორე გამოცემა (= თხზულებანი, III; ძველი ქართული ენის კათედრის შრომები, 15). თბილისი: თბილისის სახელმწიფო უნივერსიტეტი. <https://digitallibrary.tsu.ge/book/2021/july/shromebi/shanidze-qartuli-enis-gramatikis-safudzvebi-1.pdf>. – [მესამე გამოცემა] (= თხზულებანი თორმეტ ტომად, III). თბილისი: თბილისის უნივერსიტეტის გამომცემლობა, 1980. <https://archive.org/details/shanidze>.
- Topadze Gäumann (2011), Manana T. G., “The Expression of Evidentiality between Lexicon and Grammar. A Case Study from Georgian”, *Linguistic Discovery* 9, issue 2. <https://journals.dartmouth.edu/cgi-bin/WebObjects/Journals.woa/xmlpage/1/article/392>; <http://dx.doi.org/10.1349/PS1.1537-0852.A.392>.
- Topuria (1967a): ვარლამ თოფურია, სვანური ენა. I. ზმნა. (= შრომები, I). თბილისი: მეცნიერება. <https://iverieli.nplg.gov.ge/handle/1234/465825>.

- (1967b): Варлам Т. Топуриа, “Сванский язык”, in Evgenij Alekseevi Бокарев & Кетеван В. Ломтатитдзе (reds.), *Языки народов СССР. IV. Иберийско-кавказские языки*. Москва: Наука, 77–94. <https://disk.yandex.ru/i/0OQL-M6F6z-LrA>. Repr. in Topuria 2002: 68–90.
- (1985): Варлам Т. Топуриа, “Сванский язык / სვანური ენა/ The Svan Language” , იბერიულ-კავკასიური ენათმეცნიერების წელიწადიულო / *Ежегодник иберийско-кавказского языкознания / Annual of Ibero-Caucasian Linguistics* 12, 100–148. Repr. in Topuria 2002: 91–140.
- (2002): ვარლამ თოფურიძე, შრომები. II. თბილისი: ქართული ენა. <https://iverieli.nplg.gov.ge/handle/1234/465827>.

## Abbreviations

AOR	Aorist	SG	Singular
EM	Extension marker	S3	3 <sup>rd</sup> person subject
EV_PRF	Evidential Perfect	O1	1 <sup>st</sup> person object
EV_PRF_1	Evidential Perfect (prefixal element)	O3	3 <sup>rd</sup> person object
EV_PRF_2	Evidential Perfect (suffixal element)	OV	Objective version
IMPF	Imperfect	SUPV	Superessive version
MCT	Marker of continuous tense	SV	Subjective version
PRS	Present	THM	Thematic marker
PRV	Preverb		

# ქართველურ ენათა ზმნურ პარადიგმებში ევიდენციულობის მარკერის გამოყოფის თეორიული საფუძვლები და პრაქტიკული ღირებულება

მაია ლომია, ქეთევან მარგიანი

(ივანე ჯავახიშვილის სახელობის თბილისის სახელმწიფო უნივერსიტეტი)

DOI: <https://doi.org/10.62235/dk.3.2024.8517>

[maia.lomia@tsu.ge](mailto:maia.lomia@tsu.ge) || ORCID: 0009-0009-4695-7426

[ketevan.margiani@tsu.ge](mailto:ketevan.margiani@tsu.ge) || ORCID: 0009-0005-8571-3943

ევიდენციულობა უნივერსალური კატეგორიაა. იგი ყველა ენაში დასტურდება, თუმცა გამოხატვის ენობრივი საშუალებები განსხვავებული შეიძლება იყოს მონათესავე ენებშიც. ევიდენციულობა ზოგიერთ ენაში გრამატიკული კატეგორიაა, ზოგში კი მხოლოდ ლექსიკურ-სემანტიკურია. როცა ევიდენციულობა გრამატიკული კატეგორიაა, ევიდენციულობის გამოხატვის მთავარი საშუალება არის ზმნა. ევიდენციულობის ზმნური ფორმები ორგვარია: პერფექტული და იმპერფექტული. პერფექტული ზმნური ფორმები უფრო მეტად არის გავრცელებული, იმპერფექტული ფორმები კი იშვიათია და განსაკუთრებით არასალიტერატურო ენებისთვის არის დამახასიათებელი.

სალიტერატურო და არასალიტერატურო ენათა მონაცემების შესადარებლად განსაკუთრებით საინტერესო მასალას იძლევა ქართველური ენები, რადგან აერთიანებს სალიტერატურო ქართულსა და არასალიტერატურო მეგრულ, ლაზურ, სვანურ ენებს. ევიდენციულობის ზმნური ფორმები გამოხატავენ არამოდალიზებულ და მოდალიზებულ სემანტიკას.

სტატიაში განხილულია: ა) არამოდალიზებული ევიდენციულობის პერფექტული და იმპერფექტული ზმნური ფორმათწარმოებისა და სეგმენტაციის საკითხები და ბ) ევიდენციულობის მარკერის გამოყოფის თეორიული საფუძვლები.

პერფექტული ევიდენციულობის ზმნური ფორმები (ინვერსიული თურმეობითები) საერთოქართველური მოვლენაა; მათთვის ევიდენციულობის გამოხატვა მეორეული, თუმცა სინქრონიულ დონეზე გააქტიურებული ფუნქციაა. პერფექტულ ევიდენციულობის ზმნურ ფორმებს სპეციალური მორფოლოგიური მაწარმოებელი უმეტესად არ მოეპოვებათ და მხოლოდ ლექსიკურად გამოხატავენ უნახაობას; ამით უპირისპირდებიან ნახული მოქმედების გამომხატველ აორისტის ჯგუფის ზმნურ ფორმებს; ამ თვალსაზრისით სამ ქართველურ ენაში (ქართული, მეგრული, ლაზური) ერთნაირი მდგომარეობაა, ხოლო სვანური განსხვავდება მათგან; კერძოდ, სვანურში გარდამავალ უპრევერბო პერფექტულ ზმნებში უნახაობის გამოსახატავად სპეციალური -ენ სუფიქსი განვითარდა; პრევერბიან ფორმებში კი სრულად დაიხრდილა ოდენ რეზულტატიურობის გამოხატვის შესაძლებლობა და მათ მხოლოდ უნახაობის გამოხატვის უნარი შეიძინეს. ეს ფორმებიც, ცხადია, უპირისპირდება აორისტის ჯგუფის ზმნებს.

სპეციალურ ლიტერატურაში სვანური პერფექტული ევიდენციალური ზმნური პარადიგმები (მწკრივები) სხვადასხვა დასახელებით გვხვდება. ზმნურ ფორმათა სახელებების ამგვარი სიტყვით გამოწვეულია იმით, რომ სხვადასხვა მკვლევარი სხვადასხვა კრიტერიუმით ხელმძღვანელობს. ბოლო დროს განხორციელებული ევიდენციალობის კატეგორიის სისტემური კვლევისა და ქართველური მონაცემების შეჯერების საფუძველზე საანალიზო მწკრივებს ეწოდათ **ევიდენციალური პერფექტი I** და **ევიდენციალური პერფექტი II**. იმპერფექტული ევიდენციალური ზმნური ფორმები მხოლოდ არასალიტერატურო ქართველურ ენებში (მეგრული, ლაზური, სვანური) გაჩნდა; ამათგან, ლაზურში ლექსიკურადაა გამოხატული (აღწერითი წარმოებისა), ხოლო მეგრულსა და სვანურში სპეციალური მორფოლოგიური მარკერით აღინიშნება; შესაბამისად, მარკერის გამოყოფის საკითხი მხოლოდ მეგრულსა და სვანურ ენებში განიხილება.

მეგრულში იმპერფექტული ევიდენციალური ზმნური სტრუქტურული მოდელები უპირისპირდებიან ნახული მოქმედების გამომხატველ აწმყოსა და იმპერფექტის ზმნურ პარადიგმებს; აღსანიშნავია, რომ იმპერფექტული ევიდენციალური ზმნური ფორმები მეგრულში თავიდან აღწერითი წარმოების იყო; ტრანსფორმაციის შედეგად ჩამოყალიბდა ორგანული ფორმები და ევიდენციალობის გამომხატველი სპეციალური მარკერიც მაშინ გაჩნდა. მეგრულშიც აქტუალურია იმპერფექტული ევიდენციალური ზმნური ფორმების სახელებების საკითხი. ქართული სალიტერატურო ენისგან განსხვავებული მწკრივები მეგრულში პირველად გ. როგავამ გამოყო ცალკე სერიად (IV სერიად). სხვადასხვა სახელწოდების მიუხედავად, ყველა თანხმდება, რომ მათთვის საყრდენია აწმყოსა და უწყვეტლის ფუძეები და გამომხატვენი უნახაობას; ამიტომ ეწოდათ: ამ მწკრივებს **ევიდენციალური იმპერფექტი I** და **ევიდენციალური იმპერფექტი II**. მეგრულში ევიდენციალური იმპერფექტული ზმნური ორგანული ფორმები რამდენიმე ეტაპის გავლის შემდეგ ჩამოყალიბდა; კერძოდ, ნო- – -უ კონფიქსით ნაწარმოებ მიმდებარს ერთვოდა ყოფნა მეშველი ზმნა აწმყოს ან უწყვეტლის ფორმით: **ორე(ნ)** „არის“, **ორღუ** „იყო“. მეშველი ზმნები დროთა განმავლობაში გაცვდა. **ორე(ნ)** ფორმისგან მიღებული „ე“ ხმოვანი (ორე(ნ)→ ე) ნო- პრეფიქსთან დაწყვილდა და ნო- – -ე კონფიქსად გაფორმდა, როგორც ევიდენციალობის მარკერი. ამ პროცესს ხელი შეუწყო იმანაც, რომ ნო- პრეფიქსი წარსულის სემანტიკისაა; უნახაობის სემანტიკას კი წარსული დრო განსაზღვრავს.

იმპერფექტული ევიდენციალური ზმნური პარადიგმები სვანურში ქართულისაგან განსხვავებული მწკრივებია. თურმეობითი სვანური ზმნის უღლებაში განსაკუთრებით რთული მწკრივია როგორც წარმოების, ისე სემანტიკური ნიუანსების მრავალფეროვნების თვალსაზრისით. წინამდებარე სტატიაში გაზიარებულია თვალსაზრისი, რომ იმპერფექტულ ევიდენციალურ ზმნურ ფორმებს სვანურში საყრდენი ფუძისა და სემანტიკის მიხედვით ეწოდოს: **ევიდენციალური იმპერფექტი I** და **ევიდენციალური იმპერფექტი II**. იმპერფექტული ევიდენციალური ზმნური ფორმების სტრუქტურული ანალიზი სვანურში წარმოაჩენს სხვადასხვა სემანტიკის ფორმათა დაპირისპირებას; კერძოდ, ნეიტრალურ იმპერფექტს ორი ფორმა უპირისპირდება: საზედაო ქცევის **ევიდენციალური იმპერფექტი I** და **ევიდენციალური იმპერფექტი II**. საზედაო ქცევის გამომხატველ

ფორმას ემატება **-უნ-ა / -ინ-ა** სუფიქსები, რომელთა პირველი კომპონენტი (-უნ- /-ინ-) მიჩნეულია ევიდენციალობის მარკერად, რადგან უპირისპირდებიან უწყვეტელს, როგორც ნახული მოქმედების გამომხატველს. სვანურში **ევიდენციალური იმპერფექტი II** რთულია სტრუქტურულად; იგი წარმოშობით სინთეზური წარმოებისაა. ორგანულ **ევიდენციალურ ზმნურ** ფორმად ჩამოყალიბებას ხელი შეუწყო იმან, რომ ასახავდა განგრძობით მოქმედებას წარსულში. როგორც სხვა მწკრივებში, აქაც **-უნ-** ევიდენციალობის მარკერია. **-უნ** სუფიქსი არქაულია; მისი ფუნქციუს დიფერენციაცია მწკრივთა მიხედვით მოგვიანებით უნდა მომხდარიყო.

ამრიგად, ზმნური პარადიგმების სისტემური ანალიზის შედეგად გამოვლინდა, რომ ევიდენციალობის მორფოლოგიური მარკერი ქართველურ ენებს შორის მხოლოდ სვანურსა და მეგრულში დასტურდება; მათგან სვანურში მეტად არის გამოკვეთილი ევიდენციალობის გრამატიკალიზაციის პროცესი. ევიდენციალურ ზმნურ ფორმათა სეგმენტირება-გლოსირებასა და ევიდენციალობის მარკერის გამოყოფას აქვს თეორიული და პრაქტიკული ღირებულება; კერძოდ: ა. წარმოაჩენს ენის მდიდარ შესაძლებლობებს უნივერსალური კატეგორიის გამოსახატად; ბ. მიემართება ევიდენციალობის კატეგორიის გენეზისს და მის ავთენტურობას ადასტურებს; გ. გამოიყენება ტექსტების მორფოლოგიური ანოტირების დროს.



# Unknown Pages of English-Georgian Lexicography (George Ellis and his Comparative Lexicon of Kartvelian Languages)

Tinatin Margalitadze, Salome Tchighladze (Tbilisi)

(Ilia State University, Georgia)

DOI: <https://doi.org/10.62235/dk.3.2024.8518>

[tinatin.margalitadze@iliauni.edu.ge](mailto:tinatin.margalitadze@iliauni.edu.ge) || ORCID: [0000-0001-9485-1698](https://orcid.org/0000-0001-9485-1698)

[salome.tchighladze.1@iliauni.edu.ge](mailto:salome.tchighladze.1@iliauni.edu.ge) || ORCID: [0009-0003-2413-9410](https://orcid.org/0009-0003-2413-9410)

**Abstract:** In 1788 a book was published by the British author George Ellis in London under the title *Memoir of a Map of the Countries Comprehended Between the Black Sea, and the Caspian, with an Account of the Caucasian Nations and Vocabularies of their Languages*. The book describes the history, culture, and religion of the peoples of the Caucasus, and provides information about the languages spoken in the Caucasus region. While discussing the countries of the Caucasus, a certain space is dedicated to Georgia with geographic, political and ethnographic information, which is followed by a small comparative lexicon of the Kartvelian languages. It is noteworthy that the author is interested not only in the Georgian language, but also in the other Kartvelian languages, and the book contains material of Megrelian and Svan along with Georgian. It can be said that the lexicographic material included in this book is the first example of describing the material of Kartvelian languages for English readers. Our article provides short information about the book itself, and discusses Ellis's *Comparative Lexicon of Kartvelian Languages* and its sources. Special attention is paid to the transliteration rules of Kartvelian words into English, which are mainly based on the rules of Modern English spelling. Ellis's *Comparative Lexicon of the Kartvelian Languages* is important for the study of early stages in the history of English-Georgian lexicography; it is also interesting in the context of linguistic affinity, as the Kartvelian languages are presented in it exactly from this point of view.

**Keywords:** Kartvelian languages, comparative lexicon, Georgian language, Megrelian language, Svan language, transliteration rules, affinity of languages

## Introduction

On July 26, 1963, a well-known Georgian anglicist, Professor Niko Q'iasashvili, published an article in the newspaper "Literary Georgia", entitled "About One Unique Book".<sup>1</sup> Thanks to this article, the Georgian public learned for the first time about a book of the British author George Ellis, *Memoir of Map of the Countries Comprehended Between the Black Sea, and the Caspian, with an Account of the Caucasian Nations and Vocabularies of their Languages* (see Fig. 1).

The book was published in London in 1788. It describes the history, culture, and religion of the peoples of the Caucasus, and provides some information about the languages spoken in the Caucasus region. While discussing the countries of the Caucasus, a certain place is dedicated to Georgia, which is followed by a small comparative lexicon of the Kartvelian languages. It is noteworthy that the author is interested not only in the Georgian language, but also in Kartvelian languages, and the book contains material of the Megrelian and Svan languages along with

---

<sup>1</sup> Q'iasashvili 1963.

Georgian. It can be said that the lexicographic material included in this book is the first example of describing the Kartvelian languages with respect to English.

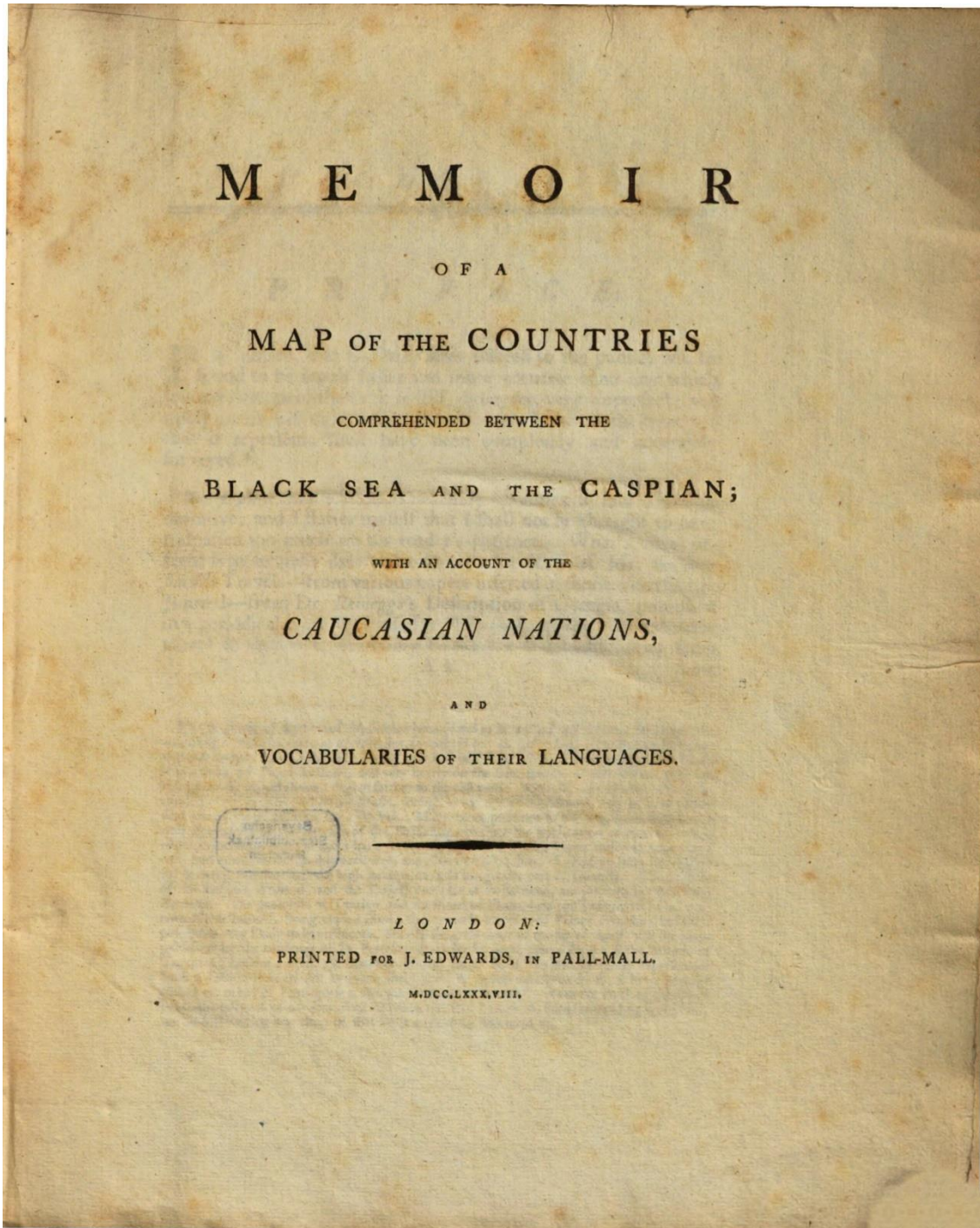


Fig. 1: George Ellis's *Memoir*, title page of the copy of Bayerische Staatsbibliothek, München

That is the reason why the book attracted the attention of Niko Q'iasashvili, along with the ethnographic, geographical, and political information it contains about Georgia. It was exactly this lexicographic material that determined our interest in this book.

We would like to extend our gratitude to Ketevan Margiani and Sophio Daraselia who helped us in the correct interpretation and analysis of the English-Svan and English-Megrelian parts of the Lexicon.

## 1. George Ellis and his Sources

As mentioned above, George Ellis's *Memoir of Map of the Countries Comprahended Between the Black Sea, and the Caspian, with an Account of the Caucasian Nations and Vocabularies of their Languages* was published in London in 1788. The book consists of three parts: 1) a map of the Caucasus (see Fig. 2); 2) short information about the peoples of the Caucasus, their history, religion, culture and some other issues; 3) small comparative lexicons of the languages of the Caucasian peoples, including the Kartvelian languages. Apart from the Kartvelian languages, Ellis's book includes glossaries of the Abkhaz, Kabardian-Circassian, Ossetian, Chechen, and Lezgian languages. As Niko Q'iasashvili points out in his above-mentioned article, the Lexicon of the Georgian language is quite adequately compiled compared to other languages, as even specialists often have difficulties in recognising words of those languages.<sup>2</sup>

George Ellis was a highly educated man who received his education at Cambridge University. He was a historian, diplomat, a member of the Parliament. In 1782–1783, Ellis was on a diplomatic mission to St. Petersburg with the British ambassador, Sir James Harris. It was during this period that he became interested in the Caucasus and the Caucasian peoples.



Fig. 2: George Ellis's Map

Ellis himself had never been to the Caucasus and, while working on his book, he relied mainly on the travel notes of Johann Anton Gldenstdt, Jacob Reineggs' description of Georgia, the works of Peter Simon Pallas and Friedrich Mller. He also used some ancient sources for his book.<sup>3</sup>

From the preface of the book we learn that George Ellis's Comparative Lexicon was based on the *Comparative Dictionary of the Languages of the Whole World* published on the order of the Russian Empress Catherine II (in Latin *Linguarum totius orbis vocabularia comparativa*, in Russian *Сравнительные словари всех языков и наречий*),<sup>4</sup> which contains material of about 200 languages of the world, including Caucasian and Kartvelian languages. For the compilation

<sup>2</sup> Q'iasashvili 1963.

<sup>3</sup> Odzeli 1998: 32–33.

<sup>4</sup> [https://archive.org/details/bub\\_gb\\_mPBLAAAAcAAJ](https://archive.org/details/bub_gb_mPBLAAAAcAAJ).



of this dictionary, Catherine II invited the famous German scientist, encyclopedist, naturalist, geographer, and traveller Peter Simon Pallas, who had been a member of the St Petersburg Academy of Sciences since 1767. The dictionary was published by the Russian Academy of Sciences in two volumes in St Petersburg in 1787–1789. According to the findings of the Georgian lexicographer Aleksandre Ghlonti, the material of the Kartvelian languages was included in the above-mentioned dictionary from the notes of Johann Güldenstädt. Ghlonti also investigated that N. Akhverdov, a Georgian prince by origin, assisted Johann Güldenstädt in collecting material of the Kartvelian languages, as he did not know the Georgian language. While working on the Kartvelian languages, the editorial team of the Comparative Dictionary also used materials provided by the Georgian scholar Anton Murav'ov.<sup>5</sup>

In the preface to his book, Ellis regretfully notes that the dictionary does not fully provide information about some languages or dialects, since this information was missing from the Russian dictionary used as a source. He also points out that the sound systems of the Caucasian languages could not be rendered accurately by English letters. These sounds were first distorted in the Russian edition due to the absence of corresponding Russian sounds, and when transliterated into the English alphabet, they became even more distant from the sounds of Caucasian languages, including Kartvelian.

In some cases, the terms “language” and “dialect” are confused in Ellis’s Lexicon. The Svan language is considered to be a dialect, while Megrelian is referred to as the Imeretian dialect. This mistake was made in the Russian dictionary, where Megrelian words are considered to belong to Imeretian dialect (*Имеретинский*) and eventually it was also introduced in Ellis’s Lexicon. Generally, the errors characteristic of the Russian edition are also found in the English version.

Despite the errors, this is a very important book, the lexicographic part of which is the first source for the English-speaking world about the Caucasian and Kartvelian languages. The Georgian-Megrelian-Svan lexical material included in the dictionary of George Ellis, and especially the Megrelian and Svan material has stirred up the interest in the scientific study of the Kartvelian languages.

## 2. The Comparative Lexicon of George Ellis

### 2.1. The English-Georgian part of the Lexicon

The Georgian language is referred to in the Lexicon as a dialect (*Carduel Dialect*). This mistake can be traced back to the Comparative Dictionary published by the Russian Academy of Sciences, in which the Georgian material is considered to belong to the Kartlian dialect (*Карталински*). The English-Georgian part of the dictionary comprises 129 words, which are arranged thematically and not alphabetically. The first words of the dictionary are *God* and *Heaven*, followed by words denoting human beings and kinship terms (*father, mother, son, daughter, sister, brother, wife, husband, woman, man*, etc.); parts of the human body and facial features (*head, body, legs, stomach, face, nose, eye, eyebrow, ear, forehead, cheeks, mouth*, etc.); natural phenomena (*wind, storm, rain, hail, lightning, snow, ice*); seasons of the year (*spring, summer, autumn, winter*); periods (*day, night, morning, evening, year, time*); celestial

---

<sup>5</sup> Ghlonti 1983: 80-84.

bodies (*sun, moon, star*); surrounding geographical environment (*tree, mountain, land, river, sea, hill, shore, etc.*); senses, feelings, perception (*sight, smell, hearing, touch, feeling, pain, love*) and others (see Fig. 3). George Ellis's Lexicon follows the Russian source in terms of the arrangement of the material. This is exactly how words are presented in the Comparative Dictionary published by the St Petersburg Academy of Sciences. These thematic groups generally coincide with word lists compiled to compare languages and establish their affinity (e.g. by Gottfried Leibniz), and the creation of such comparative dictionaries served precisely the aim to establish the fact of relatedness of languages.

As noted above, the most adequately compiled part of Ellis's Comparative Dictionary of Caucasian Languages is the Georgian dictionary. As Niko Q'iasashvili points out in his above-mentioned article, the analysis of the Georgian material clearly shows that some difficult Georgian words are transliterated into English quite accurately. For example, 'Nails – Prchkheeleeby' (ფრჩხილეები), 'Fire – Tsetskhlee' (ცეცხლი), 'Marriage – Kortseeneba' (ქორწინება), 'Evening – Mtsookhry' (მწუხარი), 'Summer – Zapkhooly' (ზაფხული) and others. Even in words with transliteration errors, these errors often concern only one letter, for example in 'Gogonebba' – გაგონება (*gagoneba*, *o* for *a*); 'Gbeely' – კბილი (*k'bili*, *g* for *k*); 'Knossa' – კნოსვა (*knosva*, *s* for *v*) and others. All errors in the transliteration of words are related to the Russian source and were transferred from it into Ellis's Lexicon.

Georgian equivalents of English words are also selected quite well. There are instances when words are not transliterated correctly, but the Georgian equivalent is adequate. There are examples when an English word has two or three Georgian equivalents out of which one may not be transliterated accurately. There are few cases when transliterated words do not correspond to any Georgian word.

## 2.2. The English-Megrelian part of the Lexicon

The English-Megrelian section of George Ellis' Comparative Lexicon contains only 61 words. As mentioned above, the Megrelian language is erroneously referred to in the Lexicon as the Imeretian dialect, an error that can be traced back to the Russian source. The number of Megrelian words is the same as in the Russian source. It should be noted that in the Comparative Dictionary of the St Petersburg Academy of Sciences, many words are missing from dictionaries of other languages and dialects, not only from the Megrelian and Svan word lists. Like the English-Georgian part, the Megrelian part of Ellis's Lexicon also contains quite competently transliterated words. All transliteration errors are related to the Russian source. The Megrelian equivalents, in the majority of cases, are also quite adequate.

## 2.3. English-Svan Part of the Lexicon

The Svan part of George Ellis' Lexicon includes only 60 words, like its Russian source, and the highest number of errors and inaccuracies are found in this part. Almost 60% of the Svan words included in the Lexicon are incorrectly transliterated, and often Svan equivalents do not correspond to English words. This fact can probably be explained by the complexity of the Svan language.

[ 77 ]

GEORGIAN LANGUAGE.

	<i>Carduel Dialect.</i>	<i>Imeretian.</i>	<i>Suaneti Dialect.</i>
God	Gmerty	Horomti	Gherbet
Heaven	Tfah	Tfash	Tfah
Father	Mamma	Mooma	Moo
Mother	Dedda	Deeda	Dee
Son	Shyilly	Skooa	Yezag
Daughter	Kaly	Ozoory	Zoonak
Brother	Tfmah	Djeema	Moohkbay
Sister	Da	Datchkym	Datchoor
Husband	Kmary	Komodjy	Chafh
Wife	Tfoly	Cheely	Aikhoo
Girl	Kally	Ozoory	Soorag
Boy	Bidjaoo	Bidjaoo	Tfhkynta
Child	Kmatfvilly, Tfvilly, Krmah	Bofhy	Bobfh
Man	Kadtsy	Kodfhy	Maray
People	Khalkhee, Erny, Catfuy	Margalee	Khvaeemary
Head	Tavee, Kavee	Doodce	Tkhoom
Face	Peeraffa, Sakhay, Peeris-fakhay	Peejeefhee	
Nose	Tfkhveery	Tchkhindee	Yepkhna
Nostril	Nefto, Nefvy		
Eye	Twaly, Tvaly	Toly	Tay
Eyebrow	Tfarby		
Eyelashes	Khaltay, Erkta		
Ear	Koory	Oodjy	Shdeem
Forehead	Shoobly	Kooa	Neekba
Hair	Tma	Toma	Patoo
Cheeks	Loka, Koba		
Mouth	Peeree	Pidjee	Peel
Throat	Kharkhanto, Tkelly		
Teeth	Kbeely, Gbeely	Keebeery	Shdik
Tongue	Aina	Neena	Nin
Beard	Tfwerry, Tfwerry	Preemooly	Waray
Neck	Kifferry, Kaily	Kifferry	Kinfhkh
Shoulder	Pkary, Mkhary	Khoodjy	Mekher
Elbow	Dakvy, Tikrtah		

Hand

Fig. 3: George Ellis's Lexicon, Georgian part, beginning (p. 77)

### 3. Transliteration Rules of Words of the Kartvelian Languages into English

One of the interesting issues of George Ellis's Lexicon is the transliteration rules of words of Kartvelian languages into English. Ellis's Lexicon does not contain any explanation of the rules that he used in his Comparative Dictionary to transcribe Georgian, Megrelian, and Svan words into English. Therefore, we have described these rules by observing the material of all three languages.

It should be noted that the transliteration rules are often quite competently compiled. This fact is also noted by Niko Q'iasashvili. As examples, he cites such difficult-to-transliterate words as 'Nakvertskhaly' (ნაკვერცხალი), 'Tseleetsadee' (წელიწადი), 'Khelmtsepaiba' (ხელმწიფეობა), 'Mtsookhry' (მწუხრი), 'Ghelva' (ღელვა), which are rendered into English with great accuracy.<sup>6</sup> For conveying the sounds of the Kartvelian languages into English, Ellis mainly relies on the Modern English spelling rules. The application of spelling rules of a target language in transliterating source language words was a common practice at that time. This is evident by the analysis of bilingual dictionaries of Georgian with Italian and Dutch of the 17<sup>th</sup>–18<sup>th</sup> centuries.<sup>7</sup> The transliteration rules discussed below apply to Georgian as well as Megrelian and Svan.

### 3.1. Vowels

1) The Georgian back vowel ა [a] is conveyed into English by the letter *a*.

2) The Georgian back vowel უ [u] is rendered into English by means of *oo* or *u*. *Oo* is pronounced in English as a long vowel [u:] (e.g. in *moon* [mu:n]). This rule of reading appeared in the Modern English period as a result of the Great Vowel Shift, when the Middle English long vowel [o:] changed into long [u:].<sup>8</sup> Consequently, the Georgian equivalent of 'forehead', შუბლი (*šubli*), is transliterated in Ellis' Lexicon as *Shoobly*.

3) The Georgian front vowel ი [i] is transcribed into English by the following vowels and vowel sequences: *ee*; *i*; *y*; *ey*; *uy*.

a) *Ee* is pronounced as a long vowel [i:] in English, e.g. in *tree* [tri:]. This rule of reading is also connected with the Great Vowel Shift, when the Middle English long [e:] changed into long [i:]. As a result, the Georgian equivalent of 'mouth', პირი (*p'iri*), is transcribed in Ellis' Lexicon as *Peeree*.

b) The transliteration of the Georgian vowel ი [i] by means of the English vowel *i* is also explained by English rules of reading, given that 'i' is pronounced as a short vowel [ɪ] in closed syllables, e.g. in *bit* [bɪt]. We can observe this transliteration rule in the word *Kinnooly* - ყინული ('ice').

c) The latter example (*Kinnooly*) shows the third rule of transliterating the Georgian vowel ი [i] into English, namely by means of *y*. The application of English *y* for the rendition of the Georgian vowel ი [i] is mostly found at the end of a word. The use of *y* instead of *i* at the end of a word has an interesting explanation in the history of the English language. This rule developed in the Middle English period and is associated with the technique of writing on expensive parchment. To save space, when working on manuscripts, scribes would write words without space between them, and the use of *y* instead of *i* marked the end of a word. We observe this rule in English words like *boy*, *toy*, etc.<sup>9</sup> Examples from Ellis' Lexicon are *Kmary* - ქმარი [kmari] 'husband'; *Tvaly* - თვალი [tvali] 'eye'.

<sup>6</sup> Q'iasashvili 1963.

<sup>7</sup> Uturgaidze 1999: 31–37; Witsen 2013: 166–167.

<sup>8</sup> Rastorgueva 2003: 200–203.

<sup>9</sup> Rastorgueva 2003: 184–186.

d) The application of the vowel sequences *ey* or *uy* for conveying the Georgian vowel *o* [i] is very rare and cannot be explained by English spelling rules.

4) Like the Georgian vowel *o* [i], there are various rules for transliterating the vowel *ჟ* [e] into English: *e*; *ay*; *ey*; *y*; *ai*.

a) As the study of the material has revealed, the most common way to transcribe the Georgian, Megrelian and Svan vowel *ჟ* [e] into English is the English letter *e*, which corresponds to the English rule of reading, given that this English letter is pronounced as [e] in closed syllables, e.g. in *red* [red], *bed* [bed].

b) The vowel sequences *ey*, *ay* are used quite often for all three Kartvelian languages in the transliteration of the vowel *ჟ* [e]. It is noteworthy that these vowel sequences denote the Georgian vowel *ჟ* [e] predominantly at the end of a word. This transliteration rule can be found in the following examples from Ellis's Lexicon: *Sakhay* – სახჟ (*saxe* 'face'); *Khay* – ხჟ (*xe* 'tree'); *Gamey* – ღამჟ (*game* 'night').

c) Cases of the transliteration of the Georgian vowel *ჟ* [e] into English by means of *y* or *ai* are rare in the Lexicon and cannot be explained by English spelling rules.

5) The English letter *o* is used to transliterate the Georgian back vowel *ო* [o] into English, which again fully complies with the English spelling rules. The English vowel *o* is pronounced as a short [ɒ] in closed syllables, e.g. in *hot* [hɒt], *cot* [kɒt], *pot* [pɒt].

#### 4.2. Consonants

1) Ellis's Lexicon is quite consistent in the transliteration of sonorants. The Georgian consonants ლ, მ, ნ, რ are consistently transliterated into English with the corresponding English consonants *l*, *m*, *n*, *r*.

2) The English fricative *s* is used to convey the Georgian fricatives ს [s] and ზ [z]. A confusion of these two Georgian consonants can be seen in the examples *Seesmaree* – სიზმარო (*sizmari* 'dream'); *Seeskhly* – სისხლი (*siskhli* 'blood').

3) The English voiceless stop *t* conveys both the Georgian aspirated stop თ [t] and the glottalised stop ტ [tʰ]. This confusion is due to the fact that there is no letter corresponding to თ [t] in the Russian language. Therefore, the Russian source conveyed both Georgian consonants თ [t] and ტ [tʰ] in the same way. This was transferred into the English Lexicon where both Georgian consonants are rendered by means of English *t*. Examples: *Tkeveely* – ტკივილი (*t'k'ivili* 'pain'); *Tovlee* – თოვლი (*tovli* 'snow').

4) The same reason explains the uniform transliteration rule for the Georgian aspirated stop ქ [k] and the glottalised stop კ [kʰ]. The two English letters *k* and *c* are used to transcribe these consonants into English. The letter *c* has two-fold reading in English: if it is followed by a back vowel, it is pronounced as a voiceless stop [k] (e.g. in *car* [ka:]), and if it is followed by a front vowel, it renders a fricative sound [s] (e.g. in *ice* [aɪs]).

5) For the transliteration of the Georgian fricative ვ [v], Ellis uses the English letters *v*; *u*; *w*; *b*.

a) The transliteration of the Georgian consonant ვ [v] with the English back vowel *u* can be traced back to the following fact from the history of the English language. The voiced fricative



[v] and the back vowel [u] were not distinguished in English spelling until almost the 17<sup>th</sup> century. This shortcoming was only regulated by the orthographic reform of the 17<sup>th</sup> century.<sup>10</sup>

b) The Georgian consonants ბ [b] and ვ [v] are sometimes confused in the Lexicon. This manifests itself in the fact that in several words in the dictionary, the Georgian voiced stop ბ [b] is transliterated by the English *v* while the Georgian fricative ვ [v] is rendered into English by the English *b*.

6) The Georgian aspirated stop ფ [p] is conveyed into English by the letters *p* and *f*. There is no sound in Georgian which would be an equivalent of the English fricative [f]. This error can be traced back to the Russian source, where there is no equivalent of the aspirated stop ფ [p]. This was transferred to the English version where the letter for the fricative *f* appeared in transliterating this Georgian consonant into English.

7) The situation is more complicated with the transliteration rules of Georgian consonants that are unknown to both Russian and English. Accordingly, those sounds that were already distorted in the Russian source became even more distant from Georgian sounds in their English rendering. Despite this fact, it should be noted that the rules of transliteration of Georgian affricates are quite competently compiled, for example with ც – *ts*; ძ – *dz*; ჩ – *tch*, *ch*. The same can be said about the transliteration rules of the Georgian back fricatives into English: ლ – *gh*, ზ – *kh*. However, these rules are not consistently implemented in the Lexicon and the above mentioned consonants are also conveyed by other letter clusters (see the table of transliteration rules in Fig. 4).

8) The Lexicon of George Ellis confuses the Georgian consonant triplets ძ [dz], ც [ts], წ [ts’], as well as the letters ჩ [ch] and ჭ [ch’], which are often transliterated into English by the same letter combinations.

9) The Georgian glottalised stop ყ [q] is unknown to both English and Russian. Therefore, the author seems to have equated it with the fricative ბ [kh] and conveys it into English accordingly, with *k* and *kh*.

Fig. 4 presents a combined table of the transliteration rules of the Kartvelian languages used in George Ellis’s Lexicon.

## 5. Conclusion

George Ellis’ Lexicon of the Kartvelian Languages is undoubtedly an interesting work for the study of the early stages in the history of English-Georgian lexicography. As noted, this dictionary is the first example of describing material of the Kartvelian languages with respect to English and it has truly awakened the interest into these languages. In the 19<sup>th</sup> century, English authors continued to study the Kartvelian languages. In 1883, Demetrius Rudolph Peacock published his Dictionary of five, as he calls them, West Caucasian languages (*Original Vocabularies of Five West Caucasian Languages compiled on the spot by Mr. Peacock*): Georgian, Megrelian, Laz, Svan and Abkhaz.<sup>11</sup> During the same period, Oliver and Marjory

---

<sup>10</sup> Rastorgueva 2003: 184–186.

<sup>11</sup> Odzeli 1998: 33–34; Kikvidze, Pachulia 2019: 15–22.

Wardrop become interested in Georgia and the Kartvelian languages <sup>12</sup> and Oliver Wardrop compiles “An English-Svan Dictionary”.

ა	ან	a	ს	სან	s; z
ბ	ბან	b; v	ტ	ტარ	t
გ	გან	g; gh	უ	უნ	oo; u
დ	დონ	d	ფ	ფარ	p; f
ე	ენ	e; ay; ey; (ai; y)	ქ	ქან	k; c
ვ	ვინ	v; w; u; b	ღ	ღან	g; gh; (h)
ზ	ზენ	s; z	ყ	ყარ	k; kh
თ	თან	t	შ	შინ	sh
ი	ინ	i; ee; y; ey; (uy)	ჩ	ჩინ	tch; ch; tsh
კ	კან	k; c	ც	ცან	ts; dts
ლ	ლას	l	ძ	ძილ	ts; tsz; dz
მ	მან	m	წ	წილ	ts
ნ	ნარ	n	ჭ	ჭარ	ch, tsh
ო	ონ	o	ხ	ხან	kh; gh; k; hk
პ	პარ	p	ჯ	ჯან	dj
რ	რაე	r	ჰ	ჰაე	h

Fig. 4: Transliteration rules in George Ellis’s Lexicon

George Ellis’ Lexicon is also interesting in the context of the study of affinity of languages, as the Kartvelian languages appear in it exactly under these considerations. The epochal works of Franz Bopp (*On the Conjugation System of Sanskrit in comparison with that of Greek, Latin, Persian and Germanic*) and Rasmus Rask (*Introduction to the Grammar of the Icelandic and other Ancient Northern Languages*), which marked the beginning of the scientific study of Indo-European languages and the development of the historical-comparative method, had not yet been published. However, by that time there was already a fairly mature linguistic view that similarities between European languages could not be accidental and might indicate their origin from some common source.

It was this interest that gave the impetus to the comparative study of languages. Various lists of words were created and scholars wrote down equivalents of these words in different languages of the world. Empress Catherine II, who commissioned the St Petersburg Academy of Sciences to create a comparative dictionary of the world’s languages, compiled the first list herself and invited Peter Simon Pallas to implement this project. It is worth noting that the thematic groups that were included in such lists (words denoting human beings and kinship terms, body parts and facial features, celestial bodies, natural phenomena, surrounding geographical environment, etc.) would later play an important role in the study of language families.

<sup>12</sup> Odzeli 1998: 34-35.

Despite this great interest in the issue of the affinity of languages, we observe methodological mistakes in terms of selecting and comparing the linguistic material. The error of this method is that authors compare words with similar meanings in different languages. This was precisely what the historical-comparative method which gave rise to the scientific study of languages has “corrected”. The change that this method has brought about is the comparison of words with similar forms and not similar meanings in different languages. Cognate words do not always have the same meaning in related languages and can undergo certain semantic changes. For example, the Kartvelian word for ‘fence’, ღობე, has preserved this meaning in Georgian and Megrelian, while it underwent a semantic change in the Svan language where it denotes a ‘beehive’.<sup>13</sup> The English word *fowl* has the meaning of a ‘domestic animal’, while its cognate in German, *Vogel*, means ‘bird’. Thus, if we compare German *Vogel* with English *bird* and not English *fowl*, the methodological mistake will be obvious. Such examples can be cited ad infinitum.<sup>14</sup>

Comparative dictionaries were created at the early stage of the development of linguistics and despite methodological errors, they played an important role in the discoveries that led to the scientific study of languages and the establishment of the historical-comparative method.

## References

- Ellis (1788): George E., *Memoir of Map of the Countries Comprahended Between the Black Sea, and the Caspian, with an Account of the Caucasian Nations and Vocabularies of their Languages*. London: John Edwards. <https://books.google.ge/books?id=MnlNAAAACAAJ>; <https://www.digitale-sammlungen.de/de/view/bsb10691490>.
- Fähnrich & Sardschweladse (1990): Heinz F., Surab S., *Etymologisches Wörterbuch der Kartwel-Sprachen*. Leiden: Brill.
- Gamkrelidze & Ivanov (1994–1995): Thomas V. G., Vjaceslav V. I., *Indo-European and the Indo-Europeans: A Reconstruction and Historical Analysis of a Proto-language and a Proto-culture*. Vols. I–II. Berlin / New York: Mouton de Gruyter.
- Ghlonti (1983). ალექსანდრე ღლონტი, ქართული ლექსიკოგრაფიის საკითხები [*Issues of Georgian Lexicography*], თბილისი: საბჭოთა საქართველო.
- Kikvidze & Pachulia (2019): Zaal K., Levan P., “Demetrius Rudolph Peacock and the Languages of Georgia”, in *General and Specialist Translation / Interpretation: Theory, Methods, Practice*. Proceedings of the International Conference, Kyiv: Agrar Media Group, 15–22. <https://www.academia.edu/38781081>.
- Odzeli (1998): მარიკა ოძელი, ქართულ-ინგლისური ლიტერატურული ურთიერთობის ისტორიისათვის [*Towards the History of Georgian-English Literary Contacts*], თბილისი: თბილისის უნივერსიტეტის გამომცემლობა.
- Q’iasashvili (1963): ნიკო ყიასაშვილი, “ერთი უნიკალური წიგნის შესახებ [About One Unique Book]”, ლიტერატურული საქართველო [*Literary Georgia*], July 26.

---

<sup>13</sup> Fähnrich / Sardschweladse 1990: 513.

<sup>14</sup> Gamkrelidze / Ivanov 1995.

- Rastorgueva (2003): Т. А. Расторгуева, *История английского языка*. Второе издание, стереотипное [A *History of English*. 2nd edition], Москва: Астрель.
- Uturgaidze (1999): თედო უთურგაიძე, ქართული ენის შესწავლის ისტორია. ნაწილი I: ქართული ენის გრამატიკა [History of the Study of the Georgian Language. Part I: Grammar of the Georgian Language], თბილისი: ქართული ენა. <https://iverieli.nplg.gov.ge/handle/1234/406118>.
- Witsen (2013): ნიკოლას ვიტსენი, ჩრდილოეთი და აღმოსავლეთი თათრეთი. საქართველო ანუ იბერია და სამეგრელო [Northern and Eastern Tartary. Georgia or Iberia and Samegrelo], თბილისი: უნივერსალი [Georgian translation of Nicolaes Wi., *Noord en Oost Tartarye*, Amsterdam: Halma 1705. <http://resolver.sub.uni-goettingen.de/purl?PPN340052120>].

# ინგლისურ-ქართული ლექსიკოგრაფიის უცნობი ფურცლები (ჯორჯ ელისის ქართველური ენების შედარებითი ლექსიკონი)

თინათინ მარგალიტაძე, სალომე ჭიღლაძე (თბილისი)

(ილიას სახელმწიფო უნივერსიტეტი)

DOI: <https://doi.org/10.62235/dk.3.2024.8518>

[tinatin.margalitadze@iliauni.edu.ge](mailto:tinatin.margalitadze@iliauni.edu.ge) || ORCID: 0000-0001-9485-1698

[salome.tchigladze.1@iliauni.edu.ge](mailto:salome.tchigladze.1@iliauni.edu.ge) || ORCID: 0009-0003-2413-9410

1788 წელს ლონდონში გამოიცა ბრიტანელი ავტორის ჯორჯ ელისის წიგნი რუკა შავი და კასპიის ზღვებს შორის მოთავსებული ქვეყნებისა, კავკასიის ხალხთა აღწერითა და მათი ენების სიტყვარებით (Ellis 1788).<sup>15</sup> მასში აღწერილია კავკასიის ხალხთა ისტორია, კულტურა, რელიგია და მოცემულია ინფორმაცია კავკასიის რეგიონში გავრცელებული ენების შესახებ. კავკასიის ქვეყნებს შორის გარკვეული ადგილი ეთმობა საქართველოს, რომელსაც ავტორი ურთავს ქართველური ენების მომცრო შედარებით ლექსიკონს. საყურადღებოა, რომ ავტორი ინტერესდება არა მხოლოდ ქართული ენით, არამედ ქართველური ენებით, და წიგნში ქართულთან ერთად წარმოდგენილია მეგრული და სვანური ენების მასალაც. შეიძლება ითქვას, რომ ამ წიგნში შესული ლექსიკოგრაფიული მასალა არის პირველი ნიმუში ქართველური ენების განხილვისა ინგლისურ ენასთან მიმართებით. სწორედ ამ ლექსიკოგრაფიულმა მასალამ განსაზღვრა ჩვენი ინტერესიც ჯორჯ ელისის აღნიშნული წიგნისადმი.

ჯორჯ ელისი უადრესად განსწავლული პიროვნება იყო, რომელსაც განათლება კემბრიჯის უნივერსიტეტში ჰქონდა მიღებული. იგი იყო ისტორიკოსი, დიპლომატი, პარლამენტის წევრი. 1782–1783 წლებში ჯ. ელისი დიპლომატიური მისიით იმყოფებოდა სანკტ-პეტერბურგში დიდი ბრიტანეთის ელჩთან, სერ ჯეიმს ჰარისთან ერთად. სწორედ ამ პერიოდში დაინტერესდა იგი კავკასიით და კავკასიელი ხალხებით. თავად ელისი არასოდეს ყოფილა კავკასიაში და, აღნიშნულ წიგნზე მუშაობისას, იგი ძირითადად დაყრდნობს ი. გიულდენშტედტის მოგზაურობის ჩანაწერებს, ასევე ი. რაინეგის საქართველოს აღწერას, პ. ს. პალასისა და ფრ. მიულერის შრომებს, ასევე ნაშრომისთვის გამოიყენა ანტიკური წყაროები (ყიასაშვილი 1963; ოძელი 1998: 32–33).

როგორც წიგნის წინასიტყვაობიდან ირკვევა, ჯორჯ ელისის შედარებითი ლექსიკონი დაფუძნდა რუსეთის იმპერატრიცა ეკატერინე მეორის დაავალებით გამოცემულ მთელი მსოფლიოს ენების შედარებით ლექსიკონს (*Linguarum totius orbis vocabularia comparativa*, რუსულად *Сравнительные словари всех языков и наречий*),<sup>16</sup> რომელშიც შევიდა მსოფლიოს 200-მდე ენის, მათ შორის კავკასიური და ქართველური ენების, მასალა. ლექსიკონი პეტერბურგის მეცნიერებათა აკადემიამ გამოსცა პეტერბურგში ორ ტომად 1787–1789 წლებში.

ლექსიკონის ინგლისურ-ქართულ ნაწილში შესულია 129 სიტყვა, რომლებიც დალაგებულია თემატურად და არა ანბანურ რიგზე. ლექსიკონის პირველი სიტყვებია: ღმერთი და ზეცა, შემდეგ კი მათ მოსდევს ადამიანისა და ნათესაობის აღმნიშვნელი

<sup>15</sup> *Memoir of Map of the Countries Comprahended Between the Black Sea, and the Caspian, with an Account of the Caucasian Nations and Vocabularies of their Languages.*

<sup>16</sup> [https://archive.org/details/bub\\_gb\\_mPBLAAAACAAJ](https://archive.org/details/bub_gb_mPBLAAAACAAJ).

სიტყვები (მამა, დედა, ვაჟიშვილი, ქალიშვილი, და, ძმა, ცოლი, ქმარი, ქალი, კაცი და სხვა); ადამიანის სხეულის ნაწილებისა და სახის ნაკეთების აღმნიშვნელი სიტყვები (თავი, ტანი, ფეხები, მუცელი, სახე, ცხვირი, თვალი, წარბი, ყური, შუბლი, ლოყები, პირი და სხვა); ბუნებრივი მოვლენების აღმნიშვნელი სიტყვები (ქარი, ქარიშხალი, წვიმა, სეტყვა, ელვა, თოვლი, ყინული); წელიწადის დროები (გაზაფხული, ზაფხული, შემოდგომა, ზამთარი); დროის მონაკვეთის აღმნიშვნელი სიტყვები (დღე, ღამე, დილა, საღამო, წელი, დრო); ციური სხეულები (მზე, მთვარე, ვარსკვლავი); გარემომცველი გეოგრაფიული გარემოს აღმნიშვნელი სიტყვები (ხე, მთა, მიწა, მდინარე, ზღვა, ბორცვი, ნაპირი და სხვა); გრძნობა-აღქმის აღმნიშვნელი სიტყვები (მხედველობა, ყნოსვა, გაგონება, შეხება, გრძნობა, ტკივილი, სიყვარული) და სხვა.

ქართული ენის მასალა ლექსიკონში მოხსენიებულია, როგორც დიალექტი (*Carduel Dialect*). ეს შეცდომა მომდინარეობს რუსეთის მეცნიერებათა აკადემიის მიერ გამოცემული შედარებითი ლექსიკონიდან, რომელიც ჯორჯ ელისის ლექსიკონის წყაროა და რომელშიც ქართული ენის მასალა მოხსენიებულია როგორც ქართული დიალექტი (*Карталински*).

უნდა აღინიშნოს, რომ ჯ. ელისის კავკასიური ენების შედარებითი ლექსიკონის ყველაზე კვალიფიციურად შედგენილი ნაწილი ქართული ლექსიკონია. ქართული მასალის ანალიზი ნათლად აჩვენებს, თუ როგორი სიზუსტითაა გადატანილი ინგლისურ ენაზე საკმაოდ რთული ქართული სიტყვები. მაგ.: Nails – Prchkheelebby (ფრჩხილები), Fire – Tsetskhlee (ცეცხლი), Marriage – Kortseeneba (ქორწინება), Evening – Mtsookhry (მწუხრი), Summer – Zapkhooly (ზაფხული) და სხვა. იმ სიტყვებშიც კი, სადაც ტრანსლიტერაციის შეცდომებია, ხშირად შეცდომა მხოლოდ ერთ ასოს შეეხება. მაგ.: Gogonebba – გაგონება (o > a); Gbeely – კბილი (G > K); Knossa – ყნოსვა (s > v) და სხვა. სიტყვების ტრანსლიტერაციის ყველა შეცდომა რუსულ წყაროს უკავშირდება და რუსული წყაროდანაა შეცდომით გადმოტანილი ელისის ლექსიკონში.

ლექსიკონის ინგლისურ-მეგრული ნაწილი მხოლოდ 61 სიტყვას მოიცავს. მეგრული ენა ლექსიკონში მოხსენიებულია, როგორც იმერული დიალექტი, რაც ასევე რუსული წყაროდანაა შეცდომით გადმოტანილი. მეგრულ ნაწილში, ისევე როგორც ინგლისურ-ქართულ ნაწილში, არის საკმაოდ კვალიფიციურად ტრანსლიტერირებული სიტყვები. ტრანსლიტერაციის ყველა შეცდომა რუსულ წყაროს უკავშირდება.

ლექსიკონის სვანური ნაწილი მხოლოდ 60 სიტყვას მოიცავს, ისევე როგორც მისი რუსული წყარო და ყველაზე მეტი შეცდომა და უზუსტობა სწორედ ამ ნაწილში ვლინდება. ლექსიკონში შეტანილი სვანური სიტყვების თითქმის 60% არასწორადაა ტრანსლიტერირებული და, ხშირად, სვანური ეკვივალენტი არ შეესაბამება ინგლისურ სიტყვას. ეს ფაქტი ალბათ სვანური ენის სირთულით უნდა აიხსნას. სვანური ენა ლექსიკონში ასევე დიალექტადაა მოხსენიებული.

ჯ. ელისის ლექსიკონის ერთ-ერთი საინტერესო საკითხი არის ქართველური ენების სიტყვების ინგლისურად ტრანსლიტერაციის წესები. ამისათვის ავტორი იყენებს ახალი ინგლისური ენის კითხვის წესებს და ისე გადააქვს ინგლისურ ენაზე ქართული, მეგრული და სვანური სიტყვები.

ჯორჯ ელისის ქართველური ენების ლექსიკონი მნიშვნელოვანია ინგლისურ-ქართული ლექსიკოგრაფიის ისტორიის ადრეული ეტაპის შესასწავლად. ეს ლექსიკონი საინტერესოა იმის გამოც, რომ მასში ქართველური ენები შეტანილია ენათა გენეზისური კავშირების კვლევისა და ენათა ნათესაობის დადგენის კონტექსტში.