A Corpus-Based Dictionary for the Endangered Megrelian Language

Irina Lobzhanidze[1], Rusudan Gersamia[1]

[1] Ilia State University, Kakutsa Cholokashvili Ave 3/5, Tbilisi 0179, Georgia
E-mail: irina_lobzhanidze@iliauni.edu.ge, rgersamia@iliauni.edu.ge

Abstract

This paper presents a corpus-based approach to compiling a bilingual Megrelian-English online dictionary. The Megrelian language belongs to the UNESCO Atlas of the World's Languages in Danger group of "increasingly endangered" languages, and faces a number of critical challenges, among them a lack of standardised resources, intergenerational transmission, and minimal digital presence. To address these gaps, we developed an annotated corpus of contemporary Megrelian, consisting of 97691 tokens and 60959 types. It is based on data collected through fieldwork in Samegrelo, Georgia, from the years 2022 to 2025. The bilingual Megrelian-English dictionaries were developed in parallel, using the same dataset processed in Fieldworks Language Explorer (FLEx, 2024). This approach enabled the integration of corpus annotations into the dictionary entries. We used lexeme-based and root-based configurations, resulting in the creation of two online dictionaries, available online. The first dictionary is oriented toward the translation of individual words, while the second focuses on the translation of individual morphemes. In the first case, each lexical entry is supported by morphosyntactic information, phonetic transcription (IPA), glosses, and semantic descriptions. In the second case, the entries represent individual morphemes, providing not only glosses, but also information about their occurrences and links to their use in the corpus. The finalised data is available online through https://xmf.iliauni.edu.ge/.

Keywords: Endangered language documentation; corpus-based lexicography; Megrelian

1. Introduction

The Kartvelian language family, native to the southern Caucasus, comprises Georgian, Megrelian, Laz, and Svan, and shares a relatively uniform sound system. In addition, it boasts a well-developed system of word inflection and derivation, and agglutinating and inflecting systems that make use not only of a large variety of grammatical affixes, but also of ablaut and other process types typical of internal stem inflection and split ergativity of the sentence.

Georgian, the most widely spoken Kartvelian language, has a rich literary tradition that began in the fifth century (Chikobava, 2008 [1952]; Shanidze, 1976, and others) and serves as the official language of Georgia. Svan, spoken in the mountainous regions of north-western Georgia, reveals significant phonological and morphological differences from other Kartvelian languages. Megrelian, spoken in western Georgia, and Laz, spoken in north-eastern Turkey, share a close relationship and display remarkable

similarities in terms of vocabulary and grammar.

Svan, Megrelian, and Laz, each of which transmit unique cultural knowledge, are classified as "increasingly endangered" in the UNESCO Atlas of the World's Languages in Danger (2021). Protecting these languages is vital for preserving oral tradition, historical memory and cultural identity. It necessitates more than language documentation, requiring preparation of dictionaries from data gathered during fieldworks.

To address the challenges faced by the Kartvelian languages, this paper is focused on the Megrelian language, drawing on data gathered through language documentation and dictionary development efforts.

Megrelian (ISO 639-3: xmf) is subdivided into two dialects: Zugdidi-Samurzakhano (ZS) and Senaki-Martvili (SM). Both are spoken across eight municipalities of the Samegrelo-Zemo Svaneti region: Abasha, Senaki, Martvili, Zugdidi, Tsalenjikha, Chkhorotsku, Khobi, and Poti. The Samurzakhano dialect is also used by the Gali population, including those who remained in Abkhazia after the war, as well as those displaced to other regions of Georgia or abroad. Megrelian is also spoken by communities in Tbilisi and by those displaced to other parts of the country.

Unlike widely spoken languages equipped with pretrained models and various linguistic tools, "increasingly endangered" languages like Megrelian lack even basic NLP tools such as annotated corpora, PoS taggers, and morphological analysers. Moreover, the complexity of their grammar and phonology require special approaches that cannot simply be adapted from high-resource languages.

Compiling the Megrelian Language Corpus (MLC) and the Megrelian-English dictionaries represents an attempt to document, analyse, and preserve this endangered language, while also enhancing its accessibility on a global scale.

This work was carried out during the language documentation project funded by the Rustaveli National Science Foundation (FR-21-993-3, 2021-2025), which aimed to collect contemporary Megrelian data through fieldwork and to process it using Fieldwork Language Explorer (FLEx). The resulting annotated data includes 97691 tokens (60959 types), and serves as the foundation for developing the online corpus, sketch grammar[1] and online dictionaries.

By combining contemporary fieldwork conducted in 2022-2024 with technological tools and traditional lexicographic methods, the project resulted in two distinct dictionaries

[1] In this paper, the term “sketch grammar” is used in two related senses. First, it refers to a brief grammatical description within the language documentation process, produced alongside corpora and dictionaries and outlining key features such as phonology, morphology, syntax and word classes (Mosel, 2006). Second, it refers to the Grammar Sketch tool in FLEx, which automatically compiles grammatical information from the annotated corpus (SIL International, 2025).

that reflect the current usage of the Megrelian language: a morphosyntactic dictionary with analytical glosses, which is technical and aimed at linguists, and a bilingual dictionary, which offers full lexical equivalents, accessible to learners and speakers.

The corpus and the dictionaries integrated with the MLC are freely available at https://xmf.iliauni.edu.ge under the Creative Commons Attribution ShareAlike 4.0 International License (CC BY-NC-SA 4.0). This resource is intended not only as a comprehensive reference for the contemporary state of Megrelian, but also as a practical tool to support its revitalisation.

The paper is subdivided into several parts: 1. *Introduction,* outlining the significance of Megrelian as part of the Kartvelian language family and introduces the project dedicated to the documentation of the Megrelian language; 2. *Background and Data Collection*, providing overviews the existing Megrelian dictionaries and represents the data collection stages; 3. *Annotation and Corpus Development,* describing the data annotation and processing stages and giving information on corpus size, linguistic coverage, etc.; 4. *The Dictionaries - Design and Generation,* presenting the configurations for both the lexeme-based and morpheme-based dictionaries, and also thoroughly describing the export and converstion stages, oulining the linkage between the corpus and the dictionary entries, and; 5. *Conclusions*, *Challenges and Future Works*, which summarises the corpus-based lexicographic approach to the Megrelian language, provides a short description of the ongoing challenges, and describes future plans concerning the use and potential improvement of the data.

2. Background and Data Collection

The first records of Megrelian appeared in the 17th century, soon after Georgia opened up to European travelers. Franciscan and Catholic missionaries, the most notable of who was, Arcangelo Lamberti, produced the first descriptions, hymn translations, and ethnographic accounts of Megrelian life (Lamberti, 1654). In the 18th century, explorers Güldenstädt (1787–1791) and Julius von Klaproth (2012 [1812–1814]) expanded this foundation by compiling Megrelian wordlists and linguistic notes, while 19th-20th century scholars (Tsagareli, 1880; Kipshidze, 1914; Javakhisvhili, 1937) started the first systematic studies of Megrelian and influenced the works of Soviet-period scholars (Chikobava, 1930, 1936; Jgenti, 1953, 1960 and others), who published monographs on Megrelian phonology, case alignment, verbal morphology and its comparison with other Kartvelian languages.

Since the 1980s, research has covered a variety of topics, among them a linguistic analysis of Megrelian and comparative analysis of the Kartvelian languages (Machavariani, 2002; Danelia & Dundua 2006; Kartozia, 2008; Kartozia et al., 2010; and others).

In parallel, scholars have generated documentation of Megrelian vocabulary, especially,

in addition to the early Megrelian wordlists of the 17th and the 18th centuries mentioned above, Erckert (1895) published a 30-language vocabulary that included Megrelian data in the Caucasian languages series, and Kipshidze (1914) released the first Mingrelian-Russian lexicon. Later works include the Chan-Megrelian-Georgian dictionary (Chikobava, 1938), a special glossary by Kilanava (2010), billingual dictionaries by Eliava (1997) and Charaia (1997), and major bilingual dictionaries such as the four-volume Megrelian-Georgian dictionary (Kajaia, 2000-2009) and the Megrelian-German dictionary (Fähnrich & Kajaia, 2001). Recent online editions (Kajaia, 2000-2009; Kobalia, 2010, 2020; and others) tend to digitise this earlier material to reflect historical rather than contemporary usage of the Megrelian language. It is worth mentioning that none of the abovementioned resources is a bilingual Megrelian-English online dictionary. These resources underline how the language's endangerment today presents complex challenges, coming as a result of both linguistic and sociocultural factors, in particular:

1. **The influence of Georgian and globalisation processes**: Increased globalisation and the dominance of Georgian have diminished the value of maintaining Megrelian. As a result, its use and cultural significance have steadily declined. Several factors provoke this loss: on the one hand, the general processes of globalisation and the influence of Georgian have undermined the importance of an unwritten spoken language; while on the other, the dispersion of Megrelian-speaking communities, caused by socio-economic and political migration over the past thirty years, has weakened transmission. As a result, the everyday language of Megrelian children and young adults has been replaced by Georgian.

2. **Urgency of preservation**: With each passing generation, the number of proficient speakers of Megrelian decreases, making the urgency of preservation critical. The younger generations do not sufficiently acquire the endangered Megrelian language due to societal and educational influences, leading to a significant gap in generational transmission between the linguistic heritage of older generations and the linguistic proficiency of the younger population. This implies that the above data collected a century ago fails to represent the current grammatical structure and vocabulary of the language.

3. **Scarsity of up-to-date reliable resources**: From a contemporary perspective, the Megrelian language suffers from a shortage of materials, be it written texts, media or documented contemporary data. There are few linguistic studies that apply modern electronic technologies to such data. The absence of a properly annotated, up-to-date Megrelian corpus compilcates understanding of the language's grammatical structure and vocabulary. At present, the only resource available online, apart from the MLC, is the Megrelian section of the Georgian National Corpus (GNC Megrelian) (Gippert et al., 2011–2025), which contains just 89404 words of unannotated texts collected in the 20th century. These limitations significantly affect the compilation of contemporary grammars,

textbooks, and dictionaries that could support both scholarship and the revitalisation of the Megrelian language.

Thus, compiling the Megrelian Language Corpus (MLC), together with the Megrelian-English morpheme and lexical dictionaries, represents the first effort to systematically document, analyse, and preserve this endangered language in a form accessible to both scholars and the public. By combining contemporary field recordings, modern annotation tools, and traditional lexicographic approaches, the project has resulted in the first resources that reflect the Megrelian language as it is spoken today.

2.1 Fieldwork and Data Collection

To address the above-mentioned linguistic and sociocultural challenges, we shifted our focus from reviewing the existing materials to collecting new data through fieldwork, ensuring that our corpus and dictionaries reflect the language as it is spoken today. The role of Megrelian language documentation was not to traditionally describe the language, but rather to collect data to support the further production of an online corpus and dictionary, with sketch grammar to be published afterwards.

Following the principles described in Austin (2006), Bowern (2008), and others, we sought to gather language samples across different genres and socio-cultural contexts, including everyday culture and toponyms, ceremonies, livelihood, and other aspects of Megrelian life.

The language documentation process resulted in 58 hours of finished recordings, spread across two years: 54 hours completed during the first two years, and an additional 4 hours completed during the final year. Recording duties were shared equally among the four field-workers; each responsible for primary material, plus an additional collection of recordings from displaced Megrelian speakers. Due to the limited server space, the final online corpus will cover approximately 150 short video files, each 1.5-4 minutes long, balanced between two major dialect zones and four age groups (15-30, 31-45, 46-60, 61+).

Each respondent was documented through a metadata sheet covering bio-demographic and sociolinguistic variables: initials, birth-year, gender, place of residence, dialect competence, knowledge of other languages, migration history, education, profession and family language profile. Respondents signed consent forms in Georgian and English, which granted or withheld permissions separately for (i) making recordings, and (ii) publishing and/or making the recordings available online (audio, video, or both). Speakers had the option to stop the recording at any time. Each recording session received a unique identifier (e.g., 0001, 0002, etc.) that was retained throughout the processing stage.

The data collection was subdivided into narratives and a two-block questionnaire. The narrative covered (i) traditional life and toponyms, (ii) rituals, myths and beliefs, (iii) cuisine and table discourse, and (iv) personal stories or folktales. The sociolinguistic questionnaire addressed behavior, beliefs, knowledge, attitudes, and attributes, as described in Dilman (1977), with the purpose of documenting language use, transmission and identity within the community. Keeping in mind that the main focus was to collect oral data and convert it into written form, we used high-quality recording equipment with video and audio recorded simultaneously.

3. Annotation and Corpus Development

Megrelian narratives were firstly transcribed into the International Phonetic Alphabet (IPA) using a special converter (Gersamia & Lobzhanidze, 2021), and were then uploaded to Fieldworks Language Explorer (FLEx, 2024), where each text was reviewed and parsed. Each sentence was presented in its transcribed form and accompanied by free translation into Georgian and into English. These translations were produced by native Megrelian speakers who are bilingual in Georgian and Megrelian, and by those who know English as a second language (See, Fig. 1).

```
<phrase guid="2c48dec1-69ba-4bee-bcb2-a5b91d94e31f">
  <item type="txt" lang="xmf">ანუ რე თარგამუჲი, ტეხირ მეეურს, ტეხირიშ კვარჩხან დო მარძგვან სანაპიროეეფს ხალხი
  ცხოვრეენს.  </item>
  <item type="segnum" lang="en">2.2</item>
  <words>
  <item type="gls" lang="en">Here is Targameuli; the river Tekhuri flows, people live on the left and right banks
  of the Tekhuri.</item>
  <item type="gls" lang="xmf-fonipa">anu re targɑmuji, t'ɛxir mɛɛurs, t'ɛxiriʃ k'vart͡ʃxan dɔ mard͡zgvan
  sanɑp'iroɛeps xalxi t͡sxovrɛens.</item>
  <item type="gls" lang="ka-Brai">ანუ [აქ] არის თარგამეული: ტეხური მიდის (=მიედინება), ტეხურის მარცხენა და მარჯვენა
  სანაპიროებზე ხალხი ცხოვრობს.</item>
</phrase>
```

Figure 1: Phrase level annotation

After translation, each sentence was segmented into tokens, and each token was fully annotated linguistically and accompanied by English glosses. This approach presupposed the existence of bilingual data and, by combining segmentation, annotation and translation, provided the foundation for compiling the Megrelian-English dictionaries. FLEx's generic XML presents a rich set of fields for each token (See, Fig. 2).

```
<word guid="eeae8e36-e109-41f9-8dbf-ddfb2df7e646">
  <item type="txt" lang="xmf">თარგამეულიი</item>
  <morphemes>
    <morph type="stem" guid="d7f713e8-e8cf-11d3-9764-00c04f186933">
      <item type="txt" lang="xmf">თარგამეულ</item>
      <item type="txt" lang="xmf-fonipa"></item>
      <item type="cf" lang="xmf">თარგამეულ</item>
      <item type="gls" lang="en">Targameuli</item>
      <item type="msa" lang="en">pn</item>
    </morph>
    <morph type="suffix" guid="d7f713dd-e8cf-11d3-9764-00c04f186933">
      <item type="txt" lang="xmf">-ი</item>
      <item type="txt" lang="xmf-fonipa">-</item>
      <item type="cf" lang="xmf">-ი</item>
      <item type="hn" lang="xmf">1</item>
      <item type="gls" lang="en">NOM</item>
      <item type="msa" lang="en">n:(Case)</item>
    </morph>
  </morphemes>
  <item type="gls" lang="en">Targameuli</item>
  <item type="pos" lang="en">pn</item>
</word>
```

Figure 2: Word level annotation

Because FLEx maintains an automatically generated morpheme-based lexicon, recurring stems and affixes were auto-filled, and each new token was either linked to an existing lexical entry with English glosses, or assigned a provisional one. This lexicon keeps corpus statistics current (97691 tokens and 60959 types), and allows immediate export of data for further archiving or analysis. The lexicon containing tokens with English glosses is the base for the MLC morpheme-based lexicon accessible through the corpus interface by choosing between *Texts, Lines* and *Morphemes*:

- In the *Texts* section of the corpus interface, entries are listed by Number, the Title in Georgian and English, by ELAN, and Video/Audio files. By clicking on a number, the full text is displayed, with sentences subdivided into words, and words further segmented into clickable morphemes, each linked to glosses, grammatical information and PoS tags. All sentences are accompanied by English and Georgian translations;

- In the *Lines* section of the corpus, each entry consists of a unique number, the sentence in Megrelian, and its English translation. The sentences and their translations are fully searchable. Clicking on a number takes the user to the corresponding place in the full text, allowing them to view the sentence in context;

- The *Morphemes* section contains the bilingual Megrelian-English morpheme-based dictionary, accompanied by information on each morpheme, its English gloss, grammatical features and frequency of occurrence in the corpus. By clicking on a morpheme, the user is taken to a list of all its occurrences in the *Lines* section. To view the full context in the corpus, the user can then click on the line where the morpheme appears.

3.1 Grammatical Tagging

Development of the corpus was closely connected to the linguistic information about Megrelian created during the annotation period. Each level of annotation (PoS tagging, morphosyntactic labeling etc.) links the Megrelian narratives to both the corpus-based analysis and the dictionary. By assigning each token a lemma and, consequently, providing grammatical information for its morphemes, annotation converts the text into a dataset that is uploaded to the SQL database and becomes searchable online afterwards.

PoS tagging of Megrelian texts means assigning part-of-speech labels and inserting them into the field indicating the word category. Defining this basic type of corpus annotation allows us to distinguish between nominal and verbal inflection used to provide the morphosyntactic labeling of morphemes. The PoS tagging followed the

tokenisation of the raw text, which was done automatically, since Megrelian word tokens in written text are normally delimited by white space. The only exemption to this rule were so-called "multiword" tokens (1-2).

(1) ɔ*sur-i-*ɛ
woman-NOM-AUX
'she is a woman'

(2) bʒɑ-t͡s'k'umɑ
sun-POST
'like a sun'

For the purposes of PoS tagging, the following tags were used (See, Table 1):

PoS	**Labels**	**Examples**
• Noun	n	*bɑɣɑnɑ* 'child', *ɔsuri* 'woman'
• Verb	v	*tvɑluns* 'thinks', *uʔɔrs* ' loves'
• Adjective	adj	*ut͡ʃɑ* 'black', *t͡ʃ'it͡ʃ'ɛ* 'small'
• Numeral	num	*xuti* 'five', *ʒɑrnɛt͡ʃi* 'forty'
• Pronoun	pro	*t͡ʃkimi* 'my', *mutuni* 'something'
• Conjunction	cnj	*dɔ* 'and', *nɑmdɑ* 'that'
• Particle	prt	*vɑr* 'yes', *kɔ* 'no'
• Adverb	adv	*tudɔ* 'below', *gɔɣɑ* 'yesterday'
• Adposition	adpos	*gurʃɛni* 'because', *umʃɔ* 'without'
• Interjection	inj	*ditɑ* 'oh', *vɑvɑ* 'wow'

Table 1: Parts of Speech

PoS tags were applied in two ways: (1) to define the part of speech for individual morphemes, and (2) to define the part of speech for entire tokens. The PoS tags assigned to individual morphemes are represented in the online dictionary of morphemes linked to the Megrelian Language Corpus (MLC), while the labels assigned to whole tokens are used in the Megrelian-English dictionary.

3.2 Lemmatisation

Lemmatisation is the process of identifying the base form of a word from one of its inflected variants, and generally corresponds to the vocabulary of the language. In Megrelian, lemmatisation is particularly difficult due to the extensive use of agglutinating affixes. For instance, a verb form can include seven prefixes (Gersamia, 2022) and seven suffixes, each encoding distinct grammatical features and appearing before and after the root. As a result, determining the lemma is a complex task.

For lexicographic purposes, the lemmatisation of Megrelian follows the principles set out in the Morpho-syntactic Annotation Framework (ISO/DIS 24611, 2012). According to MAF, verbal forms are normally lemmatised using the infinitive; nominal forms using the nominative singular (3); and adjectives using the positive nominative singular (4). However, Megrelian, like other Kartvelian languages, does not have a true infinitive. Instead, verbs are lemmatised either by the masdar[2], a verbal noun in the nominative singular, or by their verbal root (5).

(3) a) *ɔsur-i*
woman-SG.NOM
'woman'

b) *k'ɑt'u -Ø*
cat-SG.NOM
'cat'

(4) a) *ʧ'iʧɛ-Ø*
small-SG.NOM
'small'

b) *mɔ- ʧ'iʧ-ɛ-Ø*
dim-small-dim-SG.NOM
'smaller'

c) *mɑ- ʧ'iʧ -ɑ-Ø*
eqt-small-eqt-SG.NOM
'small like something'

d) *u- ʧ'iʧ -ɑʃ-i*
sup-small-sup-SG.NOM
'the smallest'

(5) a) *k'ɑk'-u-ɑ-Ø*
pounding-TS-MSD-SG.NOM
'pounding'

b) *dɔ-k'ɑk'-u*
prv-pounding-TS-3SGSBJ:AOR
'He/she/it pounded'

c) *k'ɑk'-un-s*
pounding-TS-3SGSBJ:PRS
'he/she/it pounds'

d) *dɔ-k'ɑk'-un-s*
prv-pounding-TS-3SGSBJ:FUT
'he/she/it will pound'

Another challenge is determining which verbal form should serve as the headword. Although the lemma is generally represented by the masdar (5), the verbal root actually appears in the eighth slot of the verbal template. As a result, verbs cannot be systematically indexed without taking into account preverbs, applicatives, and person markers, and it is difficult to establish connections between a masdar form and its verbal counterpart unless the user is familiar with the rules of Megrelian grammar. If only the masdar form is included, the dictionary lacks key verbal semantics. As noted above, the masdar serves as a convenient lemma, but because it is formally a noun, it does not always capture the full verbal semantics of the root, such as tense, aspect, and argument structure, and it does not reflect the use of preverbs, which may significantly change the meaning of a verb (6).

[2] A masdar is a type of action nominal or verbal noun, derived from a verb, which in many languages (e.g., Kartvelian, Arabic) show a mixture of verbal and nominal properties and, therefore, provide good examples of mixed categories. (Comrie & Thompson, 1985; Koptjevskaja-Tamm, 2005, and others).

(6) *rɑgɑdɑns* 'speaks / is speaking', *t͡ʃirxinuns* 'neighs (used for a horse)', *k'ɑrk'ɑlɑns* 'clucks (used for a hen)' etc.

To address this problem, the online dictionary combines two approaches, representing not only the masdar forms, but also the third person singular present form for semantically different cases. To summarise, decisions about lemmatisation have played an important role not only in the development of the corpus and morpheme lexicon, but also in compiling the Megrelian-English dictionaries and creating the accompanying sketch grammar.

3.3 Interlinear glossing

3.3.1 Nominal inflection

In Megrelian, nouns, adjectives, numerals, and pronouns share similar structural characteristics, but the number of morphological slots and their formation patterns differ. Nominal inflection is typically formed by suffixation, while the diminutive, equative and superlative degrees of adjectives are created by employing circumfixing (2). The lexical features of nominals encompass properties that determine how a word can be combined with affixes. For nouns, a primary lexical category is propriety (common vs proper). Pronouns are subdivided into personal, demonstrative, possessive, indefinite, interrogative, relative, reciprocal, negative, determinal and reflexive. Adjectives have a lexical feature of gradability, while numerals can be lexically subdivided into cardinal, ordinal and multiple. The schemes of nominal templates are as follows:

- Noun: ROOT -> Consonant epenthesis[3] -> Number -> Vowel Epenthesis -> Case -> Emphatic Vowel -> Postposition -> Focus -> Emphatic vowel -> Particles[1, 2, 3] (7)

(7) *d͡ʒimɑ-l-ɛp-iʃɔ-t-i-ɑ-vɑ-ɔ*

brother-E-PL-BEN-POST-EMPH-QUOT1-QUOT2-QUOT3

'As it was said[4], and for brothers'

[3] In Megrelian, the sequence of morphemes includes not only the morphemes themselves, but also phonetic insertions that serve phonotactic functions between morphemes, as well as enclitics and proclitics, most of which are particles. Depending on the morphological environment and phonological constraints, such insertions can involve two types of epenthesis: consonant and vowel.

[4] The expression 'as it was said' corresponds to the three consecutive quotative markers (QUOT1, QUOT2, QUOT3) in Megrelian. These markers convey evidential and reported speech, expressed as "as it was said" in the English translation for consistency.

- Numeral: ROOT -> Number -> Vowel Epenthesis -> Case -> Emphatic Vowel -> Postposition -> Focus -> Emphatic vowel -> Particles[1, 2, 3] (8)

(8) *vitɔʃkvit-iʃɔ-t-i-ɑ-vɑ-ɔ*

seventeen-E-BEN-POST-EMPH-QUOT1-QUOT2-QUOT3

'As it was said, and for seventeen'

- Pronoun: ROOT -> Number -> Vowel Epenthesis -> Case -> Postposition -> Focus -> Emphatic vowel -> Particles[1, 2, 3] -> Conjunction (9)

(9) a) *ɛtinɑ-Ø*
this-SG.NOM
'This'

b) *ɛtin-ɛp-iʃɔ-t-i-ɑ-vɑ-ɔ*
this-PL-BEN-POST-EMPH-QUOT1-QUOT2-QUOT3
'As it was said, and for those'

- Adjective: Degree -> ROOT -> Degree -> Consonant epenthesis -> Number -> Vowel epenthesis -> Case -> Emphatic vowel -> Postposition -> Focus -> Particle[1, 2, 3] -> Conjunction (10).

(10) *gindz͡ɛ-l-ɛp-iʃɔ-t-i-ɑ-vɑ-ni*

long-E-PL-BEN-POST-EMPH-QUOT1-QUOT2-CONJ

'As it was said, and for long ones that'

3.3.2 Verbal inflection

In Megrelian, the formation of verbal inflection is governed by several morphological features: TAM series (tense-aspect-mood), voice (active, middle, and passive), personality (unipersonal, bipersonal or tripersonal) and number (singular or plural). Verbs are categorised into main verbs and auxiliary verbs, while verbal nouns and verbal adjectives are considered verbal forms that share features with nominals. The scheme of the verbal template is as follows:

- Verb: Negation particle -> Affirmative particle or Aspect -> Preverb -> Aspect (Progressive) -> Evidentiality -> Subject or Object agreement -> Applicatives, voice, causation or potentialis -> ROOT -> Augment -> Voice, causation -> Thematic suffix or potentialis -> Tense&Aspect -> Subject or Object agreement -> Paradigm marker -> Mood -> Emphatic vowel -> Particles[1, 2, 3] -> Conjunction (11)

(11) *ʃɛ-tmɛ-v-xvɑd-u-d-i-t-k'ɔn-i-ɑ-vɑ-ɔ*

PRV-PROG-1SBJ-meet-TS-IMPF-1/2PM-1/2PL-COND-QUOT1-QUOT2-QUOT3

'we were realizing that maybe (he/she/it) would be as (he/she-it) said'

- Verbal Noun: Preverb -> ROOT -> Thematic suffix -> Verbal Noun's suffix -> Case -> Postposition -> Emphatic vowel -> Particles[1, 2, 3] -> Conjunction (12)

(12) *ɑk'ɔ-sɛr-u-ɑ-ʃɑ-x-i-ɑ-vɑ-ni*

PRV-night-TS-MSD-ALL-POST-EMPH-QUOT1-QUOT2-CONJ

'until nightfall as (he/she-it) said'

- Verbal Adjective: ROOT -> Degree -> Consonant epenthesis -> Number -> Vowel epenthesis -> Case -> Emphatic vowel -> Postposition -> Focus -> Particle[1, 2, 3] -> Conjunction (13).

(13) *munə-nɑ-t͡ʃ'ir-ɑ-s-i-ɑ-vɑ-ni*

PRV-PRT:PST-cut-PRT:PST-DAT-EMPH-QUOT1-QUOT2-CONJ

'that which has been cut as (he/she/it) said'

3.3.3 Functional words

Together with content words, the Megrelian language contains functional words, including conjunctions (14), particles (15-16), adverbs (17), postpositions (18) and interjections (19), that serve to structure sentences and express various semantic nuances. The majority of these functional words has an uninflected template consisting of a ROOT.

(14) *ɔndɔ*

Until then

(15) *vɑ-g-ɑ-ɣɔr-ɛn-t*

NEG-2OBJ-APPL.INDIR-lie-TS-1/2PLSBJ

'I'm not lying to you'

(16) *ɛgba* *gɔ-t͡ʃ'q'ɔr-d-u-da*

PRT PRV-fall-PASS-3SGSBJ-CONJ

'What if it broke and'

(17) *t͡ʃ'umaan-iʃɛ*

tomorrow-ABL

'from tomorrow'

(18) *t͡ʃ'ink'-ɛp-i-ʃɛni*

gremlin-PL-NOM-POST

'about gremlins'

(19) *dita*

oh

3.4 From FLEx to corpus database

After tokenisation, lemmatisation and interlinear glossing were completed in FLEx, each text was exported as 'Verifiable generic XML' (See, Figure 1 and Figure 2), converted using special Python script, and passed to a PHP loader, which performed the following four steps simultaneously: 1) Opened a UTF-8 connection to the project database; 2) Used xml_load_file() to parse the XML hierarchy (interlinear-text -> paragraphs -> phrases -> words -> morphemes); 3) For every node, it selected the tier values and inserted them into the corresponding table; 4) All FLEx globally unique identifiers (GUIDs) were written directly in the database. The result of this pipeline is the relational schema shown (See, Figure 3), where each table represents a level of the FLEx XML hierarchy and is linked via the preserved GUIDs.

To summarise, annotated tiers from FLEx were inserted into a fully normalised SQL database and, afterwards, into a searchable Megrelian corpus and morpheme dictionary.

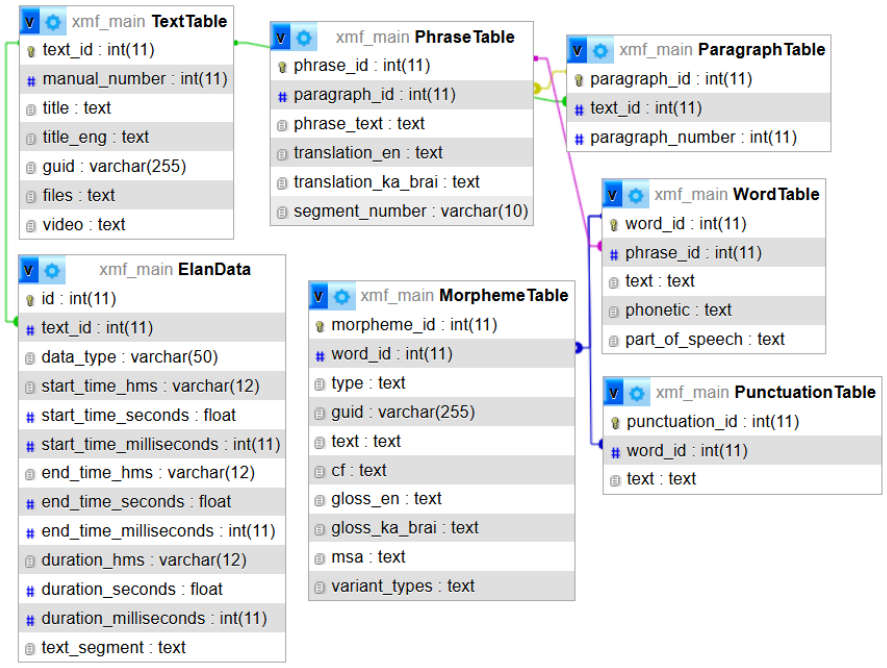

Figure 3. Corpus database

4. The Dictionaries - Design and Generation

An electronic dictionary is a database that stores lexical information and provides access to language units like separate words or MWEs, along with their senses and translations. Following Atkins & Rundell (2008), Gibbon & Van Eynde (2000) and others, four major elements can be highlighted in the development of a dictionary: a) Linguistic specification, which defines the macro- and microstructure of a dictionary; b) Technical specification, which establishes technical parameters; c) Workflow, which describes each stage of database construction, such as data input, its verification and modification (as needed); d) Mechanisms to present and disseminate lexical information to end users, which allow data access, reformatting and dissemination.

In the case of the Megrelian language, FLEx supported three of the four elements: *linguistic specification, workflow* and *technical specification*, by storing grammatical structures, managing lexical entries and providing export formats. The fourth element, *dissemination*, was achieved through custom scripts and the online portal developed within the project.

The FLEx environment was used for the representation of linguistic specification (sketch grammar used during the annotation stages), its verification and modification, and for the data conversion to make it freely accessible online. Taking into consideration that our corpus annotation consists of three main layers - transcription, morpho-syntactic, and semantic - the corpus and dictionaries' functionalities can be regarded as interconnected analytical tools. The project covered the compilation of two dictionaries:

1. A bilingual Megrelian-English dictionary (general-purpose): A lexeme-based resource containing full lexical entries (headword, IPA, part of speech, definitions and English equivalents). This is designed for learners, speakers, and general users.

2. A bilingual Megrelian-English morpheme-based dictionary (linguistic): A morpheme-based resource that lists stems, affixes, glosses, grammatical information and frequency. Entries link directly to corpus lines, making it primarily intended for linguists and researchers.

In summary, the first dictionary serves as a traditional bilingual resource with full lexical entries, while the second provides a technical tool for detailed grammatical analysis and corpus linkage.

4.1 The Bilingual Megrelian-English Dictionary (general-purpose)

The structures of dictionaries that can be generated from FLEx may come as *hybrid forms, lexeme-based*, and *root-based*. For the purpose of the Megrelian-English dictionary, we paid special attention to lexeme-based and root-based configurations. A lexeme-based configuration means that each entry is a lexeme that carries its own translation, which allows us to group under one headword different variants (20). A root-based configuration means that the entry head is the root morpheme and its derived and inflected forms appear as subentries. Such a configuration allows the researcher to find separate roots and affixes, which is important for linguistic research (21).

(20) **iprɛlnɛr** (phon. var. iprɛnɛr) *adj* any, every, of all kinds

(21) **muʃɑ-** (phon. var. muʃ-) *m* Preverb pfx PRV

As mentioned above, FLEx offers several export formats: a) Full Lexicon (lexeme-based) Standard Format Marker (SFM)[5], which exports the dictionary using Dictionary Formatter (MDF)[6] lexeme-based standard (22) and, b) Full Lexicon (root-based) SFM format, which exports the full lexicon using the MDF root-based standard. In this format, subentries are included as part of the main entry, rather than as separate entries with links to them (23).

[5] An SFM file is a plain-text format that encodes structured lexical entries, making it possible to import and export data between different linguistic software and represent dictionary and linguistic data in a structured way. Each field in this format is preceded by a backslash marker (e.g., \lx for lexeme, \ps for part of speech, \ge for English gloss) (SIL International, 2025).

[6] MDF is a standard for structuring and formatting lexical data in plain text. In its lexeme-based form, each entry is organised around the lexeme as the headword, while in its root-based form, each entry is organised around the root (SIL International, 2025).

(22) \lx **ʃxvɑdɔʃxvɑ**
\lx_xmf ʃxvɑdɔʃxvɑ
\va ʃxvɑdʃxvɑ
\vet phon. var. of
\va sxvɑdɔʃxvɑ
\vet phon. var. of
\sn 1
\ps_en Adjective
\g_en various
\sn 2
\ps_en Adjective
\g_en different

(23) \lx **ɑrtɔiɑn**
\lx_xmf ɑrtɔiɑn
\sn 1
\ps_en Adjective
\g_en combined

The lexeme-based SFM format with the .db extension was converted into .sql format and the data were made accessible through the online dictionary interface (https://xmf.iliauni.edu.ge/vocabulary). The content is visualised via the search function and presented to users in the form shown in Fig. 4:

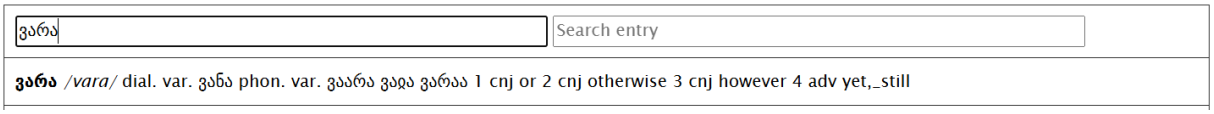

Figure 4. Dictionary entry

The user interface of the bilingual Megrelian-English general-purpose dictionary has two types of searches: (1) A lexeme-based search (Search lexeme), which allows the user to find a specific lexeme along with its IPA transcription, for reading purposes, dialectal variants, part-of-speech information and translated meanings. If a lexeme has more than one meaning, these are listed and numbered; (2) An entry-based search (Search entry), which allows the user to find information on IPA transcription, dialectal variants, parts of speech and translated meanings within an entry without focusing on the key lexeme. This approach helps the user avoid the verbal lemma problem mentioned above, and allows them to find not only keywords, but also additional information. Moreover, it enables the dictionary to be used bidirectionally.

4.2 The Bilingual Megrelian-English Morpheme-Based Dictionary (linguistic purpose)

For the purpose of creating a Megrelian morpheme-based dictionary linked to the corpus, texts with words segmented into morphemes, so-called “lexicon entries”, were exported from FLEx into the Verifiable Generic XML format. This XML file contained structured information such as <paragraph>, <phrases>, <word>, <morphemes>, and <morph> tags. Then, the file was parsed using a Python script, which mapped the tags to relational database tables (See, Fig. 3). However, only the linguistically relevant

fields (as presented in Table 2) were essential for the morpheme-based dictionary, and for facilitating integration with the website front-end.

Fields	**Examples**	
• Surface word in Mkhedruli script (UTF-8 range 10A0–10FF)	მარგაჯურო (mɑrgɑjurɔ)	
• Morpheme segmentation of the surface form	mɑrgɑjur	ɔ
• Lexical entry, including any phonetic variants	mɑrgɑlur+dial. and phon. var.	
• Lexical gloss	Megrelian	Ess
• Lexical grammatical information (case, number, etc.)	adj	adj:(case)
• Word gloss	Megrelian	
• Word category (part of speech)	adj	

Table 2: Data fields

The resulting .sql file was uploaded into the database, and the linkage between the corpus and the morpheme-based dictionary was established using FLEx's GUIDs to ensure consistent and reliable connections between the corpus data and the corresponding dictionary entries. As a result, the morpheme-based dictionary linked to the corpus is freely accessible online, see https://xmf.iliauni.edu.ge/morpheme, in the following form (See, Fig. 5):

Morpheme	Gloss	Gramm. Info	Occurences
ართმა	Search...	Search...	
ართმაჟია	each_other	rcprn	2

Figure 5. Morpheme-based dictionary entry

The user interface offers three types of searches: (1) A morpheme-based search, which allows users to locate a morpheme and, by clicking on it, view the line where the morpheme occurs in the corpus. Clicking on that line then displays the full text in context; (2) A gloss-based search, which allows users to find morphemes through their English glosses. Depending on the morpheme, the gloss may represent either its English translation or its grammatical function, e.g., all morphemes marked with ESS (essive case); (3) A grammar information-based search, which allows users to search by the grammatical category or feature expressed by a morpheme. In addition, the interface provides information about the frequency of each morpheme in the corpus.

To summarise, the vocabulary and morpheme-based search interfaces provide access to the Megrelian-English lexical database, with a vocabulary view that facilitates user-friendly browsing and a general search, while the morpheme-based search enables in-depth linguistic analysis by revealing the internal structure and usage of words.

5. Conclusions, Challenges and Future Work

This paper has detailed the development of corpus-based, bilingual Megrelian-English online lexeme-based and morpheme-based dictionaries. The underlying data was collected via fieldwork implemented in Samegrelo, Georgia, and, can be considered as a foundation for the MLC, aimed at the maintenance and preservation of the low-resourced and endangered Megrelian language. The primary goal of this project was to support language documentation by making high-quality linguistic data freely available online. As a result, using FLEx, it became possible to generate lexical datasets, export morpho-syntactically annotated texts in Verifiable Generic XML format, and transform files using custom Python scripts. This workflow allowed us to preserve grammatical information, including IPA, glosses, etc., and provide direct links to corpus examples, enriching the morpheme-based dictionary with contextual usage. Despite the work done, the following challenges remain:

a) Lemmatisation and dialectal variant representation are not always easy, especially in the lexeme-based dictionary. Megrelian's complex morphosyntax and numerous so-called “phonetically presupposed” elements, represented in templates across different PoS-es, make consistent representation challenging;

b) Constraints with user interface, which must be improved to allow wildcard search, IPA-based filtering in the morpheme dictionary, and auto-suggest options. Additionally, a Georgian-language interface is needed;

c) As the system depends on custom Python scripts developed for the project, any future changes to FLEx export schemes may require updates and technical maintenance.

The compilation of Megrelian dictionaries linked to the corpus will be useful for the further development of computational approaches to Megrelian, and possibly to other Kartvelian languages. In the next phase, the plan is to adapt our current morphological and segmentation tiers to the Universal Dependencies (de Marneffe et al. 2021) format (token, lemma, UPOS, XPOS, FEATS etc.), and use the existing annotated data as training dataset for a UDPipe (Straka et al., 2016) neural model, which will support automatic lemma prediction and morphological analysis. These automatically generated analyses will be used as draft annotations for the corpus and dictionaries' workflows.

6. Acknowledgements

This work was supported by Shota Rustaveli National Science Foundation of Georgia (SRNSFG) [grant number: FR-21-993-3, 2021-2025]. The authors would like to express their gratitude to the three anonymous reviewers for the many helpful comments and suggestions and to the organizing committee of the ELEX for their support.

7. Glosses

1SBJ	1st person singular subject
1/2PLSBJ	1st or 2nd person plural subject
1/2PL	1st or 2nd person plural
1/2PM	Person marker 1st/2nd person
2OBJ	2nd person object
3SGSBJ	3rd person singular subject
ABL	Ablative case marker
adj	Adjective
adv	Adverb
adpos	Adposition
ALL	Allative case marker
AOR	Aorist tense
APPL.INDIR	Applicative indirect
AUX	Auxiliary verb
BEN	Benefactive case marker
cnj	Conjunction
COND	Conditional
CONJ	Conjunctive marker
DAT	Dative case marker
E	Epenthesis
EMPH	Emphatic marker
FUT	Future tense
IMPF	Imperfective aspect
inj	Interjection
MSD	Masdar marker
NEG	Negation particle
n	Noun
NOM	Nominative case
num	Numeral
PASS	Passive voice
pfx	Prefix
PL	Plural
POST	Postposition
PoS	Part of Speech
PROG	Progressive aspect
PRS	Present tense
prt	Particle
PRT:PST	Past tense particle

pro	Pronoun
PRV	Preverb
QUOT1	Quotative marker 1
QUOT2	Quotative marker 2
QUOT3	Quotative marker 3
SG	Singular
TS	Thematic suffix
v	Verb

8. References

Austin, P. K. (2006). Data and language documentation. In N. P. J. Gippert, *Essentials of language documentation*. Berlin & New York: Mouton de Gruyter, pp. 87-113.

Bowern, C. (2008). *Linguistic fieldwork: A practical guide*. New York: Palgrave Macmillan.

Chikobava, A. (1930). *čanuris gramatikuli analizi [Grammatical analysis of Zan]*. Tbilisi: enis, istoriisa da materialuri kulturis institutis (enimkis) moambe [Bulletin of the Institute of Language, History, and Material Culture (ENIMKI)].

Chikobava, A. (1936). *čanuris gramatikuli analizi: tekʻstebitʻa da ganmartebebitʻ [Grammatical analysis of Zan: accompanied by texts and explanations]*. Tbilisi: enis, istoriisa da materialuri kulturis institutis (enimkis) moambe [Bulletin of the Institute of Language, History, and Material Culture (ENIMKI)].

Chikobava, A. (2008 [1952]). *enatʻmecʻnierebis šesavali [Introduction to linguistics]*. Tbilisi: tʻbilisis saxelmcipʻo universiteti (Tbilisi State University).

Comrie, Bernard; Thompson, Sandra A. (1985). Lexical nominalization. In T. *Shopen, Language typology and syntactic description 3: Grammatical categories and the lexicon*. Cambridge: Cambridge University Press, pp. 349-398.

Danelia, N. & Dundua, I. (2006). megruli enis prakʻtikuli kursi [Practical course of the Megrelian language]. In R. Amirejibi-Mullen, N. Danelia, & I. Dundua, *kolxuri (megrul-lazuri) ena [Colchian (Megrelian–Laz) language]*. Tbilisi: Universal, pp. 175-339.

Dilman, D. (1978). *Mail and Telephone Surveys: The Total Design Method*. New Jersey: Wiley.

Erckert, R. (1895). *Die Sprachen des kaukasischen Stammes [The languages of Caucasian origin] (with a foreword by Prof. Friedrich Müller)*. Vienna: Hölder.

Gersamia, R. (2022). The morphonemics of verbal prefixes in Megrelian. In I. Roy, N. Boneh, D. Harbour, & O. Matushansky, *Building on Babel's rubble*. Paris 8: Presses Universitaires de Vincennes, pp. 37-59.

Gibbon, D. & Van Eynde, Fr. (2000). *Lexicon Development for Speech and Language Processing*. London: Kluwer Academic Publishers.

Güldenstädt, J. A. (1787–1791). *Reisen durch Russland und im Caucasischen Gebirge*

[Travels through Russia and in the Caucasus Mountains] (Vols. 1–2; P. S. Pallas, Ed.). St. Petersburg: Kaiserliche Akademie der Wissenschaften.

Javakhishvili, I. (1937). *Qarṭuli da kavk'asiuri enebis tavdap'irveli buneba da nat'esao ba [The original nature and kinship of Georgian and Caucasian languages].* Tbilisi: sssr mec'nierebat'a akademiis sak'art'velos p'ilialis gamomc'emloba [Publishing House of the Georgian Branch of the USSR Academy of Sciences].

Jgenti, S. (1953). *čanur-megrulis p'onetika [Phonetics of Zan and Megrelian].* Tbilisi: t'bilisis saxelmcio universitetis gamomc'emloba [Publishing House of Tbilisi State University].

Jgenti, S. (1960). *k'art'velur enat'a šedarebit'i p'onetika [Comparative phonetics of Kartvelian languages].* Tbilisi: t'bilisis saxelmcio universitetis gamomc'emloba [Publishing House of Tbilisi State University].

Kartozia, G. (2005). *lazuri ena da misi adgili k'art'velur enat'a sistemaši [The Laz language and its place within the Kartvelian language system].* Tbilisi: Nekeri.

Kartozia, G. (2008). *megruli da lazuri tek'stebi [Megrelian and Laz texts].* Tbilisi: Nekeri.

Kartozia, G., Gersamia, R., Lomia, M. & Tskhadaia, T. (2010). *megrulis lingvisturi analizi [Linguistic analysis of Megrelian].* Tbilisi: Meridiani.

Kipshidze, I. (1914). *Grammatika mingrel'skogo (iverskogo) yazyka s khrestomatiey i slovarem [Grammar of the Mingrelian (Iverian) language with chrestomathy and dictionary].* Saint Petersburg: Tipografiya Imperatorskoy Akademii Nauk [Printing House of the Imperial Academy of Sciences].

Klaproth, J. v. (2012 [1812-1814]). *Reise in den Kaukasus und nach Georgien: unternommen in den Jahren 1807 und 1808 [Journey to the Caucasus and to Georgia: undertaken in the years 1807 and 1808].* Charleston, SC [Halle]: Nabu Press [Hallisches Waisenhaus].

Koptjevskaja-Tamm, M. (2005). Action nominal constructions. In *M. S. Dryer, M. Haspelmath, B. Comrie, & D. Gil, The world atlas of language structures.* Oxford: Oxford University Press, pp. 254-257.

Lamberti, A. (1654). *Relatione della Colchide hoggi detta Mengrella, nella quale si tratta dell'origine, costumi, e cose naturali di quei paesi [Description of Colchis, today called Mingrelia, discussing the origins, customs, and natural features of those lands].* Napoli: Camillo Caualli.

Machavariani, G. (2002). *k'art'velur enat'a šedarebit'i gramatika (lek'c'iebis kursi) [[Comparative grammar of the Kartvelian languages: Course of lectures]].* Tbilisi: saxelmcip'o universitetis gamomc'emloba [State University Press].

de Marneffe, M.-C., Manning, C.D., Nivre, J. & Zeman, D. (2021). Universal Dependencies. *Computational Linguistics 47 (2)*, 255–308.

Mosel, U. (2006). Sketch grammar. In *W. Bisang, H. H. Hock, & W. Winter, Essentials of Language Documentation.* Berlin & New York: Mouton de Gruyter, pp. 301-311.

Samushia, K. (1971). *k'art'uli xalxuri poeziis masalebi (megruli tek'stebi) [Materials of Georgian Folk Poetry (Megrelian Texts)].* Tbilisi: mec'niereba [Science].

Shanidze, A. (1973). *k'art'uli gramatikis sap'użvlebi, morp'ologia (Foundations of*

Georgian Grammar, Morphology), I. Tbilisi: t‘bilisis saxelmcip‘o universiteti (Tbilisi State University).

Straka, M., Hajič, J. & Straková, J. (2016). UDPipe: Trainable Pipeline for Processing CoNLL-U Files Performing Tokenization, Morphological Analysis, POS Tagging and Parsing. *In Proceedings of the Tenth International Conference on Language Resources and Evaluation (LREC'16).* Portorož, Slovenia: European Language Resources Association (ELRA), pp. 4290–4297

Tsagareli, A. (1880). *Mingrel'skie ėtiudy, vypusk I [Megrelian studies, Part I].* Saint Petersburg: Tipografiya Imperatorskoy Akademii Nauk [Printing House of the Imperial Academy of Sciences].

Websites:

FLEx: SIL Feildworks Language Explorer (FLEx), Version 9.1. Accessed at: https://software.sil.org/fieldworks/. (20 July 2025)

GNC: Gippert, J., Meurer, P. & Tandashvili, M. (2011-2025). *The Georgian national corpus (GNC).* Accessed at: http://gnc.gov.ge/. (20 July 2025)

MC: Gersamia, R. & Lobzhanidze, I. (2022-2025). *The Megrelian Converter (MC).* Accessed at: https://xmf.iliauni.edu.ge/converter/converter.html. (20 July 2025)

MLC: Gersamia, R. & Lobzhanidze, I. (2022-2025). *The Megrelian Language Corpus (MLC).* Accessed at: https://xmf.iliauni.edu.ge/. (20 July 2025)

Dictionaries:

Charaia, P. (1997). *megrul-k‘art‘uli lek‘sikoni [Megrelian-Georgian Dictionary].* Tbilisi: SPB.

Chikobava, A. (1938). *čanur-megrul-k‘art‘uli šedarebit‘i lek‘sikoni [Chan-Megrelian-Georgian comparative dictionary].* Tbilisi: mec‘nierebat‘a akademia [Academy of Sciences].

Eliava, G. (1997). *megrul-k‘art‘uli lek‘sikoni [Megrelian-Georgian Dictionary].* Tbilisi: Intellect.

Fähnrich, Heinz; Kajaia, Otar. (2001). *Mingrelisch-Deutsches Wörterbuch (Kaukasienstudien) [Megrelian–German Dictionary (Caucasus Studies)].* Wiesbaden: Dr Ludwig Reichert Verlag.

Kajaia, O. (2000-2009). *megrul-k‘art‘uli lek‘sikoni [Megrelian-Georgian dictionary]* (Vols. 1–4). Tbilisi: Nekeri.

Kilanava, B. (2010). *900 megruli sitqva [900 Megrelian Words].* Tbilisi: Intellect.

Kipshidze, I. (1914). *Megrelian-Russian Dictionary.* St. Petersburg: Typography of the Imperial Academy of Sciences.

Kobalia, A. (2010-2020). *k‘art‘ul-megruli lek‘sikoni [Georgian-Megrelian Dictionary].* Tbilisi: Artanuji.