

## ენათმეცნიერება Linguistic

### FLEx-ის პროგრამაში მეგრული ტექსტების ანალიზის შედეგები<sup>1</sup>

რუსუდან გერსამია

ილიას სახელმწიფო უნივერსიტეტი  
ORCID: 0000-0001-7490-0851

ირინა ლობჯანიძე

ილიას სახელმწიფო უნივერსიტეტი  
ORCID:0000-0002-3040-1847

[rgersamia@iliauni.edu.ge](mailto:rgersamia@iliauni.edu.ge), [irina\\_lobzhanidze@iliauni.edu.ge](mailto:irina_lobzhanidze@iliauni.edu.ge)

<https://doi.org/10.52340/lac.2024.32.05>

**SJIF 2024: 8.268**

სამეცნიერო პროექტი „მეგრული ენის ანოტირებული კორპუსი ფორმალური გრამატიკითა და ელექტრონული ლექსიკონით“ გულისხმობს სამეგრელოში 2022-2023 წლებში ველზე მოპოვებული მასალის დამუშავებას *FieldWorks Language Explorer (FLEx, SIL International 2024)* პროგრამის გამოყენებით, რომელიც უზრუნველყოფს მონაცემთა მართვას, ონლაინ გრამატიკის, ლექსიკონისა და მონაცემის ბაზის შექმნას, რის საფუძველზეც საბოლოოდ მიიღება პროგრამიდან ავტომატურად გენერირებული მეგრულ-ინგლისური ლექსიკონი და მოკლე ფორმალური გრამატიკა.

მონაცემების ანოტირების შემდეგ ანოტირებული ტექსტები მონაცემთა ბაზის სახით განთავსდა ინტერნეტში. აღნიშნულ ბაზაში, სხვაგვარად რომ ვთქვათ, მეგრული ენის ანოტირებულ კორპუსში, განთავსდა შემდეგი ინფორმაცია: წინადადებად დაშლილი ტექსტების ქართულ და ინგლისურენოვანი თარგმანები, ტრანსკრიფცია IPA - საერთაშორისო ფონეტიკური ანბანის გამოყენებით, მორფემად დაშლილი სიტყვების ანოტაცია ლაიფცივის გლოსირების წესებსა და ევროტიპის სახელმძღვანელო პრინციპებზე დაყრდნობით, POS - მეტყველების ნაწილთა ანოტაცია.

ამ ეტაპისათვის FLEx-ის კორპუსში წარმოდგენილია 2022-2024 წლის სავსე სამუშაოებისას მოპოვებული ტექსტებიდან გადარჩეული სხვადასხვა მოცულობის 149 ტექსტი, რაც შეადგენს 9 235 წინადადებას, 97 329 ტოკენს, 30 479 უნიკალური ფორმას. მონაცემებზე მუშაობის დროს შედგენილი ლექსიკონი და მოკლე ფორმალური გრამატიკა.

მეგრულ-ინგლისური ლექსიკონი ეფუძნება FLEx-ის გარემოში შეტანილი ტექსტების მორფოსინტაქსურად ანოტირებულ მონაცემებს. FLEx-დან შესაძლებელია შეიქმნას სტრუქტურულად ლექსემასა და მორფემაზე (ძირი, აფიქსები), ან ორივეზე დაფუძნებული ლექსიკონები. მოკლე ფორმალური გრამატიკა ეფუძნება FLEx-ში სიტყვების მორფოლოგიურ სეგმენტირებას და მორფემათა კლასიფიკაციებს.

<sup>1</sup> კვლევა ხორციელდება შოთა რუსთაველის საქართველოს ეროვნული სამეცნიერო ფონდის მხარდაჭერით (გრანტი FR-21-993. პროექტი „მეგრული ენის ანოტირებული კორპუსი ფორმალური გრამატიკითა და ელექტრონული ლექსიკონით“).

**საკვანძო სიტყვები:** მეგრული ტექსტების, მონაცემის ბაზის შექმნას, ანოტირებული ტექსტები

სამეცნიერო პროექტი „მეგრული ენის ანოტირებული კორპუსი ფორმალური გრამატიკითა და ელექტრონული ლექსიკონით“ გულისხმობს სამეგრელოში 2022-2023 წლებში ველზე მოპოვებული მასალის დამუშავებას FieldWorks Language Explorer (FLEX, SIL International 2024) გამოყენებით, რომელიც უზრუნველყოფს მონაცემთა მართვას, ონლაინ გრამატიკის, ლექსიკონისა და მონაცემის ბაზის შექმნას. რის საფუძველზეც საბოლოოდ მიიღება პროგრამიდან ავტომატურად გენერირებული მეგრულ-ინგლისური ლექსიკონი და მოკლე ფორმალური გრამატიკა.

მონაცემების ანოტირების შემდეგ ანოტირებული ტექსტები მონაცემთა ბაზის სახით განთავსდა ინტერნეტში. აღნიშნულ ბაზაში, სხვაგვარად რომ ვთქვათ, მეგრული ენის ანოტირებულ კორპუსში, განთავსდა შემდეგი ინფორმაცია: წინადადებად დაშლილი ტექსტების ქართულ და ინგლისურენოვანი თარგმანები, ტრანსკრიფცია IPA - საერთაშორისო ფონეტიკური ანბანის გამოყენებით, მორფემებად დაშლილი სიტყვების ანოტაცია ლაიფციგის გლოსირების წესებსა (Lehman 2004; Comrie et al. 2008) და ევროტიპის სახელმძღვანელო პრინციპებზე (Bakker et al. 1993) დაყრდნობით, POS - მეტყველების ნაწილთა ანოტაცია (ცხრილი 1).

8.1	Word base	-	თაქ	ვგხოვრენქ				დაბადებამე
	Word ipa		tak	vtʰxɑvɾɛnk				dɑbɑdɛbɑfɛ
	Morphemes		თაქ	ვ-	გხოვრ	-ენ	-ქ	დაბადება
	Lex. Entries		თაქ	ვ- <sub>1</sub>	გხოვრ	-ენ <sub>2</sub>	-ქ <sub>2</sub>	-შე <sub>2</sub>
	Lex. Gloss		here.PROX	SUBJ1	live	TS	SUBJ1SG	დაბადება
	Lex. Gram. Info.		loc	v:(Subject&Object agreement)	L2	v:(Thematic Suffix, Potentialis)	v:(Subject&Object Agreement)	birth
	Word Cat.		loc	m				ABL
								L2
								ger:(Case)
								ger

Free en I have lived here since my birth.  
 kat აქ ვგხოვრობ დაბადებიდან.

ცხრილი 1. შიდახაზოვანი ლექსიკური და გრამატიკული ინფორმაცია FLEX-ში

ამ ეტაპისათვის FLEX-ის კორპუსში წარმოდგენილია 2022-2024 წლის საველე სამუშაოებისას მოპოვებული ტექსტებიდან გადარჩეული სხვადასხვა მოცულობის 149 ტექსტი, რაც შეადგენს 9 235 წინადადებას, 97 329 ტოკენს, 30 479 უნიკალური ფორმას. მონაცემებზე მუშაობის დროს შედგენილი ლექსიკონი და მოკლე ფორმალური გრამატიკა.

### 1. ელექტრონული ლექსიკონი

მეგრულ-ინგლისური ლექსიკონი ეფუძნება FLEX-ის გარემოში შეტანილი ტექსტების მორფოსინტაქსურად ანოტირებულ მონაცემებს. FLEX-დან შესაძლებელია შეიქმნას სტრუქტურულად ლექსემასა და მორფემაზე (ძირი, აფიქსები), ან ორივეზე დაფუძნებული ლექსიკონები.

ა) ლექსემაზე დაფუძნებულ კონფიგურაციაში მთავარ ჩანაწერებად გამოიყენება კომპლექსური ფორმები, რომლებიც წარმოადგენენ ერთ ლექსემას, ლექსიკური მნიშვნელობის მქონე ერთეულს. ამგვარი კონფიგურაცია ძირითადი მნიშვნელობების საფუძველზე ჩანაწერების ორგანიზებით ამარტივებს ნავიგაციას ლექსიკონში. სახელური პარადიგმისთვის ლემა წარმოდგენილია მხოლოდითი რიცხვის სახელობითი

ბრუნვის ფორმით, ხოლო ზმნის პარადიგმისთვის - შეიძლება წარმოდგენილი იყოს მასდარის ფორმით.

ბ) ძირეულ და აფიქსურ მორფემებზე დაფუძნებულ კონფიგურაციაში, ძირითადი ჩანაწერების სახით გამოიყენება: ძირეული ფორმები, რომლებიც მეგრულის შემთხვევაში შეიძლება იყოს მონოფონემურიც, და ფლექსიური აფიქსები, რომლებიც გვაწვდიან ინფორმაციას გრამატიკული მნიშვნელობების შესახებ. მორფემის ყველა დასაშვები ფონეტიკური ვარიანტი ან ალომორფი უკავშირდება ამოსავალ მორფემას, გაბმულ ტექსტს და მორფემათა სიებს.

## 2. მოკლე ფორმალური გრამატიკა

მოკლე ფორმალური გრამატიკა ეფუძნება FLEx-ში სიტყვების მორფოლოგიურ სეგმენტირებას და მორფემათა კლასიფიკაციებს. ცნობილია, რომ ქართველურ ენებს შორის მეგრული აგლუტინაციის ყველაზე მაღალი ინდექსის მქონეა (მელიქიშვილი 2010): სახელების შემთხვევაში მორფემები ავტონომიური და ავტომატურია, ცალსახა მიმართებაა გრამატიკულ ფორმასა და მნიშვნელობას შორის. FLEx-ში სიტყვათა მორფოლოგიური ანალიზი ცხადყოფს, რომ ზმნა, განსაკუთრებით მისი პრეფიქსული ნაწილი, მორფემათა მძლავრი ფუზიურობით ხასიათდება, რაც მორფემათა სეგმენტირებას ართულებს. არცთუ იშვიათად მორფემები რეპრეზენტირებულია არასრული ან გარდაქმნილი ფონემატური სტრუქტურით; ზოგჯერ, საერთოდ გამოტოვებულია, დაყვანილია ნულამდე, ან მოკვეცილია (მაგალითად, ზმნის ფინიტურ ნაწილში ერთი ან რამდენიმე მორფემის მოკვეცა ქმნის ე.წ. კვეცილ სტილს); ამგვარი მოდელების ანალიზისათვის FLEx-ში ხდება მათი აღდგენა ამოსავალი და/ან სრული ფორმით, რათა დაზუსტდეს გრამატიკული მნიშვნელობები და გაირჩეს ომორფემები.

სახელებისა და ზმნის მორფოსინტაქსური ელემენტების ჯაჭვი შეიძლება დაიყოს გრამატიკული მნიშვნელობის მქონე მორფემებად, კლიტიკებად (ნაწილაკები და კავშირები) და ფონოტაქტიკურად ღირებულ ჩანართებად, რომლებიც მორფემათა საზღვარზე არსებული არაკანონიკური ფონემათმომდევრობების დასაძლევად არის გამოყენებული.

სახელს მხოლოდ ორი გრამატიკული კატეგორია აქვს: ბრუნვა და რიცხვი. მორფემათა მიმდევრობა სახელების სტრუქტურაში კი გაცილებით მეტი ელემენტია: *ძირი* -> *ეპენტეზური თანხმოვანი* -> *რიცხვი* -> *ეპენტეზური ხმოვანი* -> *ბრუნვის ნიშანი* -> *ემფატიკური ხმოვანი* -> *თანდებული* -> *ფოკუსის ნაწილაკი* -> *ემფატიკური ხმოვანი* -> *ნაწილაკები* [1, 2, 3] -> *კავშირი*.

FLEx-ის ბაზა აჩვენებს, რომ მორფემათა მიმდევრობა იდენტური რჩება სხვა სახელებთან, საწყისთან და მიმდებარესთანაც. ყველაზე მეტ თავისებურებას კი ავლენენ ნაცვალსახელები; მატრიცებში განაწილებულია ალომორფები, თავისუფლად მონაცვლე ან ფონეტიკური პოზიციით შეპირობებული ვარიანტები; შეზღუდულია თანხმოვნური ეპენტეზის, რიცხვის მარკერის გამოვლენა; თანდებულთა რაოდენობა მწირია; ბრუნვის ნიშნები მრავალგვარი ფონეტიკური ვარიანტითაა წარმოდგენილი და სავსე მასალა ახალ მონაცემებსაც ავლენს. გარდა ამისა, FLEx-ის მონაცემთა ბაზა ცალსახად ავლენს ნაცვალსახელთა მნიშვნელობების მიხედვით კლასიფიცირებისა და მათ შინაარსობრივ ჯგუფებში გადანაწილების საკითხს.

მორფემათა მიმდევრობა ზმნის პრეფიქსული ნაწილისათვის ასე გამოიყურება: *უარყოფის ნაწილაკი* -> *დადასტურებითი ნაწილაკი / სრული ასპექტი* -> *ზმნისწინი* ->

უსრული ასპექტი -> თურმეობითობა -> პირის მარკერები -> ქცევა / გვარი / კაუზატივი პოტენციალისი -> ძირი.

FLEx-ში დიდი რაოდენობის მონაცემების ანალიზის შედეგად დადასტურდა აქამდე სამეცნიერო ლიტერატურაში გამოთქმული ვარაუდები, რომ ნაწილაკთა დიდი ნაწილი მოდალური ელემენტებია, შემთხვევითა ნაწილში გრამატიკალიზებულია მხოლოდ *ქო*-ნაწილაკი და მას მასპექტირებელი ფუნქცია აქვს. ცალსახად მასპექტირებელი ფუნქცია ენიჭება *გუ* (> *გ-*) პრეფიქსსაც, განსაკუთრებით მაშინ, თუ ის რთულ ზმნისწინს უსწრებს. *თმ/მ*-ნაწილაკი უსრული ასპექტისა, კერძოდ, პროგრესივის, გამომხატველია და იგი სისტემურად ვლინდება ზმნური სტრუქტურის ერთსა და იმავე პოზიციაზე სენაკურ-მარტვილურის ფორმებში. პირის ნიშნისა და ქცევის ნიშნის ადგილი არასტაბილურია, პირის ნიშანი მოძრაობს ხმოვანპრეფიქსული მორფემის (განსაკუთრებით, ქცევის ნიშნის) ან ზმნის ძირისკენ, შემდგომ კი კანონიკური ფონოტაქტიკის მიხედვით ასიმილირდება და ალომორფად იქცევა.

მორფემათა მიმდევრობა ზმნის ფინიტური ნაწილისათვის ანალიზის ამ ეტაპზე ასე გამოიყურება: *ძირი* -> *აუგმენტი* -> *გვარი* / *კაუზატივი* -> *თემატური სუფიქსი* -> *პოტენციალისის* / *პასივის*, *თურმეობის ნიშანი* -> *სავრცობი (დრო, ასპექტი)* -> *პარადიგმის ნიშანი / კავშირებითის ნიშანი* -> *პირისა და რიცხვის ნიშნები* -> *ემფატიკური ხმოვანი* -> *პირობითი კილოს მარკერები* -> *კლიტიკები [მოდალური ნაწილაკები, კავშირები]*.

ზმნის ფინიტური ნაწილი მდიდარია უღლების კატეგორიებით, ზოგჯერ მიმდევრობაში რამდენიმე თემის ნიშანია, რაც მეგრული ზმნის ფუძის შენების პრინციპებს აჩვენებს. ზმნის ფინიტური ნაწილი მოდალური ნაწილაკებისა და კავშირების ენკლიტიკებად ქცევის პარალელურად მიდრეკილია ბოლოკიდური მორფემების ჩამოკვეცისკენ ისე, რომ ზმნის გრამატიკული მნიშვნელობის გაგებას ხელი არ ეშლებს. FLEx-ის სისტემაში, ზოგადად, კარგად არის რეპრეზენტირებული გრამატიკული მნიშვნელობის გადმოსაცემად აუცილებელი და საკმარისი, ასევე ჭარბი მორფემების რაოდენობა. პროგრამა კარგად ასახავს ენის ეკონომიურობის პრინციპს მორფოლოგიის დონეზე.

FLEx-ში მეგრული ტექსტების ანალიზი საინტერესო სურათს იძლევა არაერთი თვალსაზრისით, მათ შორისაა ნასესხობების მორფოლოგიური მოდელირების და მორფომოდელში კოდების ცვლის საკითხი; მეტყველების ნაწილთა (POS) ანოტირებისას ფორმისა და ფუნქციის შეთანადების საკითხები; *რე* > *ე* „არის“, *რდუ* > *დუ* „იყო“ ზმნის ენკლიტიკით მიღებული შედგენილი შემასმენლის მესამე პირის ფორმათა სეგმენტირებისა და ანოტირების საკითხები და სხვ.

#### ლიტერატურა:

1. კარტოზია, გ., გერსამია, რ., ლომია, მ., ცხადაია, თ., - მეგრულის ლინგვისტური ანალიზი, თბილისი: გამომცემლობა „მერიდიანი“. 2010. 699 გვ.
2. მელიქიშვილი, ი., - ფლექსიურობის (ფუზიურობის) შესახებ ქართველურ ენებში. სამეცნიერო კრებული „ენათმეცნიერების საკითხები“, # I-II, გვ. 84-106. თბილისი. 2010.
3. ქირია,ჭ., ეზუგბაია, ლ., მემიშიძე, თ., ჩუხუა, მ., - ლაზურ-მეგრული გრამატიკა. I. მორფოლოგია. თბილისი: გამომცემლობა „მერიდიანი“. 2015. 900 გვ.
4. ქობალავა, ი., - გე- ზმნისწინის მნიშვნელობისათვის მეგრულში: სამეცნიერო კრებული „ენათმეცნიერების საკითხები“, 4, გვ. 166-174. თბილისი. 2002.
5. ქობალავა, ი., - აწმყოს ფუძის ერთი თავისებურებისათვის მეგრულში (გრამატიკული და სემანტიკური ასპექტები): სამეცნიერო კრებული „ენათმეცნიერების საკითხები“, #I-II, გვ. 234-251. თბილისი. 2010.

6. Comrie, B., Haspelmath, M., and Bickel, B. (2008). *The Leipzig Glossing Rules: Conventions for Interlinear Morpheme-by-morpheme Glosses*. Leipzig: Max Planck Institute for Evolutionary Anthropology.
7. International, SIL. 2024. *SIL Fieldworks Language Explorer (FLEX)*, Version 9.1. January 01. <https://software.sil.org/fieldworks/>.
8. Lehmann, Ch., 2004, "Interlinear morphemic glossing." In Booij, G., Mugdan, J., & Skopeteas, St., (eds.), *Morphologie. Ein internationales Handbuch zur Flexion und Wortbildung*. 2. Halbband. Berlin & New York: W. de Gruyter (Handbücher zur Sprach- und Kommunikationswissenschaft, 17.2). pp.1834-1857.
9. Lobzhanidze I., Gersamia R., Tsulaia N., Compiling a bilingual Megrelian-English online dictionary – Preserving endangered Kartvelian languages. In K. Š. Despot, A. Ostroški Anić, & I. Brač (Eds.), *Lexicography and semantics: Proceedings of the XXI EURALEX International Congress Cavtat, Croatia*. pp. 614-627.

**Rusudan Gersamia**

Ilia State University  
ORCID: 0000-0001-7490-0851

**Irina Lobzhanidze**

Ilia State University  
ORCID:0000-0002-3040-1847

<https://doi.org/10.52340/lac.2024.32.05>

**SJIF 2024: 8.268**

## Results of the analysis of Megrelian texts in the FLEX program

### Abstract

The scientific project "Annotated Corpus of the Megrelian Language with Formal Grammar and Electronic Dictionary" involves the processing of materials collected in Samegrelo during 2022-2023 using the FieldWorks Language Explorer (FLEX, SIL International 2024) program. This program provides data management and facilitates the creation of an online grammar, dictionary, and database, ultimately resulting in an automatically generated Megrelian-English dictionary and a brief formal grammar.

After the data annotation, the annotated texts have been made available online in the form of a database. In this database, or, in other words, in the annotated corpus of the Megrelian language, the following information is included: translations of texts split into sentences in Georgian and English, transcription using IPA (International Phonetic Alphabet), morpheme-based annotation of words based on the Leipzig Glossing Rules and Eurotyp's guiding principles, and POS (part-of-speech) annotation.

At this stage, the FLEX corpus contains 149 texts of various lengths selected from the texts obtained during fieldwork in 2022-2024, comprising 9,235 sentences, 97,329 tokens, and 30,479 unique forms. A dictionary and a brief formal grammar have been compiled during the work on the data.

The Megrelian-English dictionary is based on morphosyntactically annotated data entered into the FLEX environment. It is possible to create lexicons structured around either lexemes or morphemes (roots, affixes), or both. The brief formal grammar is based on the morphological segmentation of words and the classification of morphemes within FLEX.

**Key words:** Megrelian texts, database, data annotation

რეცენზენტი: პროფესორი თამარ ლომთაძე