

Flex (Field Works Language Explorer) პროგრამის გამოყენება მეგრული ტექსტებისთვის

წულაია ნინო

სხულუხია თამუნა

ილიას სახელმწიფო უნივერსიტეტი, თბილისი

DOI: <https://doi.org/10.52340/idw.2023.43>

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

აბსტრაქტი. თანმედროვე ტექნოლოგიები აქტიურად გამოიყენება ჰუმანიტარულ დარგებში. აუცილებელი გახდა მისი გამოყენება ენათმეცნიერებაშიც. ტექნოლოგიების განვითარების კვალდაკვალ შესაძლებელი გახდა ენობრივი ბაზების შექმნა, რამაც ენის დიაქრონული და სინქრონული ჭრილით წარმოჩენას და შენარჩუნებას შეუწყო ხელი. აქტუალური გახდა ენობრივი კორპუსების შექმნა, რამაც შესაძლებელი გახდა ენა შენარჩუნდეს იმავე სახით, როგორც ფუნქციონირებს ის მოცემულ მომენტში. ენის კორპუსები განსაკუთრებით დროულია უმწერლობო, გადაშენების საფრთხის წინაშე მყოფი ენებისთვის, რომლებიც გვხვდება მთელი მსოფლიოს მასშტაბით და ასევე საქართველოში. ერთ-ერთი ასეთია მეგრული, ის უმწერლობო ენაა, რომელზეც შეიძლება ვთქვათ, რომ დაკარგვის საფრთხის წინაშეა. ჩვენი სამომავლო მიზანია სწორედ მეგრული ენის კორპუსის შექმნას უკავშირდება, სადაც განთავსდება როგორც ადრეული, აგრეთვე თანამედროვე, სამეგრელოს სხვადასხვა რაიონში სამუშაო ველზე მოპოვებული მონაცემები. მისი შექმნის შედეგად შესაძლებელი გახდება მეგრულის ფართო ჭრილში შესწავლა და გამოკვლევა. სოციოკულტურული და ლინგვისტური (ფონოლოგიური, მორფოლოგიური) ტენდენციების გამოვლენა.

დროის ცვილებასთან ერთად იცვლება ენის გრამატიკული საკითხები, მნიშვნელოვანია მათი დაფიქსირება და გადაწყვეტა. მიზანშეწონილია კორპუსის ლექსიკოგრაფიული თვალსაზრისით გამოყენება. არსებული პროექტის ფარგლებში კორპუსის შექმნასთან ერთად ჩვენი მიზანია ელექტრონული და ბეჭდური ლექსიკონის შექმნა. განსაზღვრული სამუშაოები ხორციელდება პროგრამა FLEx-ის (Field Works Language Explorer) გამოყენებით. ველზე მოპოვებული მასალა მუშავდება აღნიშნული პროგრამის საშუალებით. მთლიანი სამუშაო პროცესის ძირითადი ეტაპებია: მიზნობრივი მასალის მოპოვება და ჩაწერა სამუშაო ველზე; მოპოვებული მასალის აღწერა და პონექტიკური გარჩევა; მონაცემების განთავსება პროგრამაში და მისი მორფოლოგიური ანალიზი. აღნიშნული კორპუსი ხელს შეუწყობს ენის დაცვას; მნიშვნელოვნად დაეხმარება ლინგვისტებსა და მკვლევრებს.

საკვანძო სიტყვები: მეგრული ტექსტები. მორფოლოგიური ანოტირება FLEx-ში. მეგრული ენის კორპუსი.

შესავალი

სხვადასხვა სამეცნიერო დარგებში და მათ შორის ენათმეცნიერებაში ძალიან აქტუალური და საჭირო გახდა სხვადასხვა ენებზე არსებული ტექსტების დიგიტალიზაცია, რაც, ცხადია, ხელს უწყობს როგორც მათ შენარჩუნებასა და დაცვას, ასევე აიოლებს ამ ენების სხვადასხვა მიმართულებით შესწავლას, დამუშავებას, კვლევას და სხვ. თანამედროვეობის კიდევ ერთი გამოწვევა არის კორპუსული ლინგვისტიკა, რომელიც დღითიდღე უფრო მოთხოვნადი და მნიშვნელოვანი ხდება. ის იძლევა საშუალებას გადავარჩინოთ ენები, მათ შორის უმწერლობო და გაქრობის საფრთხის წინაშე მყოფი,

სწორედ შენარჩუნებას და მის ფართო ასპექტში შესწავლას ემსახურება მეგრული ენის ანოტირებული კორპუსის შექმნა ფორმალური გრამატიკითა და ელექტრონული ლექსიკონით. პროცესის უზნრუველყოფა ხდება პროგრამა FLEx-ის საშუალებით. პროექტი მიზნად ისახავს მეგრული ტექსტების მოპოვებას ველზე, აღნიშნულ პროგრამაში განთავსებას და ანოტირებას.

FLEx (Field Works Language Explorer) პროგრამის აღწერა

FLEx (Field Works Language Explorer) არის პროგრამა, რომელიც გამოიყენება ენობრივი და სოციოკულტურული მონაცემების მართვისთვის, ლექსიკური და ტექსტური მასალის დამუშავებისთვის. პროგრამა, ძირითადად, განკუთვნილია საველე და კორპუსული ენათმეცნიერებისთვის. ის გამოიყენება მონაცემების მორფოლოგიური, ფონოლოგიური, ლექსიკოგრაფიული, სოციოლინგვისტური და ანთროპოლოგიური პრობლემების გადასაჭრელად.

პროგრამა შედგება ხუთი ძირითადი ნაწილისგან:

ლექსიკონის ნაწილი

ეს კომპონენტი შეიცავს ფორმებს (ფუძეებს, ძირებს, მორფემებს, ფრაზებს და ა.შ.) და მათ აღწერილობებს.

პროგრამის მნიშვნელოვანი ნაწილია ლექსიკონის ველი, სადაც თავმოყრილია ყველა ფორმა მთლიანი მონაცემიდან. ფორმები დალაგებულია ანბანურად. მოცემული მონაცემის საფუძველზე შესაძლებელია როგორც ელექტრონული ასევე ბეჭდური ლექსიკონი შექმნა, რაც ზემოთ აღნიშნული პროექტის ერთ-ერთი სამომავლო მიზანია.

ტექსტების და სიტყვების ნაწილი

შეიცავს ტექსტებს და ტექსტებში მოცემულ სიტყავათა ჩამონათვალს. ტექსტების გამოყენება შესაძლებელია სიტყვების გასანალიზებლად. ველზე მოპოვებული სრული მასალა სწორედ ამ ნაწილშია თავმოყრილი.

გრამატიკის ნაწილი

ეს პუნქტი განკუთვნილია მორფოლოგიური ანალიზისთვის, სადაც შეიძლება განისაზღვროს კონკრეტული გრამატიკული წესები, რომლებიც არეგულირებს პროგრამის მონაცემებს.

ე.წ. „ნოუთბუქის“ ნაწილი

ე.წ. ნოუთბუქი ენობრივი და კულტურული მონაცემების მართვის საშუალებას იძლევა. ეს ინსტრუმენტი ხელს უწყობს მონაცემების შესახებ დასკვნების დოკუმენტირებას, კატეგორიზაციას, ანალიზსა და შეჯამებას.

ჩამონათვალის/სიის ნაწილი

სიის არეალი არის მონაცემების შინაარსის თავმოყრის ადგილი და ძირითადად მონაცემების ჟანრს ასახავს. სიები კონფიგურირებული და მორგებულია საკვლევი საკითხის საჭიროებებს.

ველზე მოსაპოვებელი მასალის შერჩევა და განთავსება FLEx-ში

მეგრული ენა სამეგრელო-ზემო სვანეთის რეგიონშია გავრცელებული და ძირითადად ორ - ზუგდიდურ-სამურზაყანოულ და სენაკურ-მარტვილურ დიალექტებადაა წარმოდგენილი. სავალე მასალის მოპოვებისათვის პირველ ეტაპზე შეირჩა სამეგრელოს სამი რაიონი - მარტვილი, ჩხოროწყუ და წალენჯიხა, სადაც განხორციელდა მეგრული ენის მონაცემების მოპოვება. წინასწარ განისაზღვრა რესპონდენტების ასაკი, სქესი და თემატიკა. ამ ეტაპისათვის პროგრამაში მეგრულენოვან ინფორმატორთა 22 საათიანი ტექსტური მასალა, 49 ტექსტია წარმოდგენილი. არსებული მონაცემების თემატიკა ძირითადად მეგრულთა ყოფა-ცხოვრებას, რწმენა-წარმოდგენებს, წეს-ჩვეულებებს, ქორწილისა და დაკრძალვის რუტუალებს, სასოფლო მეურნეობასა და სხვა კულტურულ ასპექტებს ეხება.

ველზე მპოვებული მონაცემების ჩაწერა ვიდეოაპარატებითა და ხმის ჩამწერი მოწყობილობებით განხორციელდა. ტექსტების უშუალოდ სამუშაო პროგრამაში განთავსებამდე, საწყის ეტაპზე შეირჩა სამიზნე თემატიკის მასალები, განხორციელდა მათი აკუსტიკური გაშიფრვა თითოეული ფონემის დონეზე.

პროგრამა FLEx-ის (Field Works Language Explorer) აღწერა

ტექსტებისა და სიტყვების ველში გამოიყოფა შემდეგი პუნქტები: Info; Baseline; Gloss; Analyze, Tagging; Print view; Text Chart.

ჩვენი ამჟამინდელი საორიენტაციო პუნქტია Analyze, სადაც ერთი კონკრეტული თემატიკის მასალაა მოცემული გასაანალიზებლად. მთლიანი ტექსტი დაყოფილია აზნაცებად, წინადადებებად და სიტყვებად. წინადადებების გამყოფ საზღვრებად მიჩნეულია წერტილი. აზნაცები ინომრება ავტომატურად მათი გამოყოფის შესაბამისად. თითოეული წინადადება თარგმნილია ქართულად, ხოლო შემდეგ ინგლისურად შესაბამის ველებში.

თითოეული გამოყოფილი სიტყვა გაანალიზებულია შემდეგი პუნქტების მიხედვით:

1. word base - ამ ნაწილში მოცემულია თითო მთლიანი სიტყვა მორფემების გამოყოფის გარეშე.

2. Word ipa - თითოეული სიტყვა მოცემულია საერთაშორისო ანაბანი IPA-თი.

3. Morphemes - დასახული მიზნებიდან გამომდინარე ყველაზე მნიშველოვანი ველია, იგი პასუხისმგებელია სიტყვების გლოსირებაზე - მორფემებად დაშლაზე საჭიროების

შესაბამისად. სახელების შემთხვევაში შეიძლება გამოიყოს სიტყვის ფუძე, შემდგომ კი მასთან დაკავშირებული - ბრუნვის¹, მრავლობითობის აღმნიშვნელი, თანდებულების (-ნერი, - წკუმა), კავშირ-ენკლიტიკები (-ნ /-ნი) სხვათა სიტყვის ან შეკითხვის აღმნიშვნელი სუფიქსები. ზედსართავი სახელის შემთხვევაში ზედსართავის მაწარმოებელი ხარისხების აფიქსები. ასევე ნულოვანი (ძირითადად ბრუნვის შემთხვევაში) ან გამოტოვებული მორფემები.

ზმნების შემთხვევაში კი გამოიყოფა ზმნის დაუშლელი ძირის² ირგვლივ არსებული მორფემები: ზმნისწინების ფართო და მრავალფეროვანი ფორმები; ქცევის, მწკრივის, თემის, პოტენციალისის, ზმნის პირისა და რიცხვის ნიშნები.

პროგრამაში წინასწარაა შეტანილი სავარაუდო მორფემები, რომლებიც შეიძლება გამოიყოს სახელების ან ზმნის გლოსირების შემთხვევაში. მორფემების გამოყოფა შესაძლებელია ტირეთი, მისი დასმისას ავტომატურად გამოიყოფა აფიქსები, რომლებიც წინასწარაა დასწავლილი პროგრამის მიერ.

3. Lex. Entries, Lex. Gloss, ძირითადად ორიენტირებულია საჭირო მორფემის შემოთავაზებაზე. აქ გვხვდება საჭირო ან სავარაუდო მორფემების ჩამონათვალი, რომელთაგან ვირჩევთ ერთ რომელიმე კონკრეტულს. ეს პუნქტი ასევე უზრუნველყოფს ფორმის ან მორფემის აღწერას, შესაძლებელია მიეთითოს გამოყოფილი ერთეული ფორმის რა ნაწილია (ფუძე, პრეფიქსი, ძირი ...); კატეგორია (ფუძის ან მორფემის კატეგორია); მისი ინგლისური ეკვივალენტი. ამ პუნქტში ასევე შესაძლებელია მორფემის ვარიანტის (დიალექტური ვარიანტი) ან ალომორფების მითითება.

4. Lex. Gram. Info. მნიშვნელოვანი ველია, სადაც ნაჩვენებია მთლიანი დაუგლოსავი ერთეულის ზოგადი კატეგორია. შესაბამის ნიშნულზე დაჭერისას შემოთავაზებულია სახელების და ზმნების შემადგენელი ნაწილები ცალ-ცალკე. კატეგორიის არჩევა შესაძლებელია საჭიროების შესაბამისად.

Flex-ში მუშაობისას გამოვლენილი შედეგები

1. გრძელი ხმოვნები

გრძელი ხმოვნები ძირითადად სენაკურ-მარტვილურ დიალექტში დასტურდება. გრძელი ხმოვნები შეიძლება შეგვხვდეს სახელსა და ზმნაშიც, სიტყვის დასაწყისში, შუაში ან დასასრულს, ის შეიძლება იყოს ნებისმიერი ხმოვანი. შეიძლება მოცემული იყოს რამდენიმე

¹ ბრუნვებთან აღსანიშნავია: აბლატივის (-იმე/-შე - დაშორებითი) და ალატივის (-იმა / -შა - მიმართულებითი), დანიშნულებითის (-იმო(თ)/-შო(თ))

² მეგრული ენა აგლუტინაციური ენების რიგს მიკვკუთვნება, რომელშიც ზმნის ძირის გარშემო მრავალი გრამატიკული კატეგორიაა წარმოდგენილი. ფუძის გარშემო მორფემათა რაოდენობა შესაძლოა 10-ზე მეტიც იყოს. ზმნის მთავარ ბირთვის ძირი წარმოადგენს. (Kobalava, Gersamia 2020; Gersamia 2021)

სიტყვაში თანმიმდევრულად:

(1) ეჩღომქვითი სერით **დიფშაყარულთ** არმა ახადეფი, პაკრიშკემს ქიმიიბლანთ, დაჩხირს ქოდვოროძანთ. „ოცდაშვიდში, ღამით შევიკრიბებით ახალგაზრდები, მოვიტანთ ბორბლებს, ცეცხლს დავანთებთ“.

(2) გოდას მიღუდესია **ქვანეფი**. „მთაზე გვქონდაო ყანები“.

(3) მიკააქუნდეეს ტრადიციული **მიკაქუნაიეფი** „ეცვათ ტრადიციული ტანსაცმელი“.

ბოლო მაგალითში გრძელი ხმოვანი იკავებს სახელობითი ბრუნვის ადგილს და ითავსებს მის ფუნქციას.

გარდა სახელობითი ბრუნვისა, გრძელი ხმოვანი გვხვდება სხვა ბრუნვის ნიშნებთან, მაგალითად დაშორებით ბრუნვასთან:

(4) ძირითად **სამარგადოშეე**, მუთ უმრავლესწილო, მუთ გოოჩაანს ბინეხიინ, თაურეშე მუთმოხვადუუ მომარაგებაა, ჩქინ კუთხუში. „სამეგრელოში ძირითადად, რაც კი ვენახებია, [ყველას] მომარაგება აქედან ხდებოდა, ჩვენი კუთხიდან“.

სენაკურ-მარტვილურში გრძელი ხმოვნები დასტურდება მრავლობითი რიცხვის ფორმებთანაც:

(5) დიხაში **პრობლემეეფ** რე ჩქინ სოფეეს, ვარე დიდ მაშტაბიშ დიხეფი. „მიწის პრობლემებია ჩვენს სოფელში, არ არის დიდი ზომის მიწები“.

2. იოტის საკითხი

მუშაობის პროცესში იოტა აღინიშნა ორივე დიალექტში. ის ზოგჯერ ხმოვანთან გვხვდება, ზოგჯერ კი /ლ/ სონორის ალომორფად მარტვილურ-სენაკურში. იოტას პოზიციები დადგენილია, ის შეიძლება შეგვხვდეს როცა 1. ზმნას ერთვის ზმნისწინი ან ეს ზმნისწინი გართულებულია ნაწილაკით და მის შემადგენლობაში არის **გი** ფონემური მიმდევრობა. 2. **გე** ზმნისწინის ვოკალური დიფთონგიზაცია მომდევნოდ მდგომმა **ო** ფონემამ შეიძლება განაპირობოს (კარტოზია და სხვ. 2010; გერსამია, 2009)

(6) **გიეფშითჯ** ტეხირიშა ტანიშ ობონუშა ართოო. „ერთად დავდიოდით ტეხურაში საბანაოდ“.

(7) ათეს **ეოფშანდ** ჩაით. „ამას აავსებდა ჩაით“.

(8) ჯორიშ ნაი თორს, **კატუე** ბირცხად თოს, ფუი თაშ გააქარი. „ჯორის ნალი თვალში კატის ბრჭყალები თვალში, ასე გაქრეს“.

(9) მაჟარნეჩა დღაას სტოფის გინმედვაფუედ, **საკურთხი** სტოფს. „მეორმოცე დღეს იდებოდა საკურთხის მაგიდაზე“.

მაგალითი (6)-ში იოტა /ლ/ სონორის ალომორფია. მაგალითი (7)-ში კი მისი ფუნქცია აშკარად ფონეტიკურია.

(8) და (9) მაგალითებში დამატებით პოზიციას ავლენს და ის ნათესაობითი ბრუნვის ფორმითაა წარმოდგენილი.

3. სუბიექტური პირის ნიშნები

ხშირია **გ-** პირველი სუბიექტური პირის მონაცვლედ წარმოდგენილი **ბ-,პ-,ფ-**

ალომორფების გამოყენების შემთხვევები, რაც თავის დროზე გუდავა-გამყრელიძის მიერ დადგენილი კანონზომიერებაა (გუდავა, გამყრელიძე, 1981). ჩვენი ტექსტების მიხედვით აღნიშნული ალომორფების თავისუფალ მონაცვლეობას სენაკურ-მარტვილურში ადასტურებს გ. კარტოზიას და სხვების ნაშრომი (კარტოზია და სხვ. 2010)

(10) თაქ **გილ-ვ-ოხეთჳ** დო ქიფჩიებუთ მუდგარენ-მუდგარენეფს. “აქ ვსხედვართ და ვლაპარაკობთ რაღაცებს“.

(11) ჭიე-ჭიეთ **ქომო-ბ-რთით** ჩხოროწყეშა „ნელ-ნელა მოვედით ჩხოროწყემდე“.

(12) **პა გელა-პ-ტახია** „ტოტი მოვტეხეო“.

დასტურდება ისეთი შემთხვევები, როცა ზმნის ძირი გახლეჩილია და სუბიექტური პირი შუაშია მოქცეული.

(13) დო ჩქე იმ დიხას **დო-ფო-ბ-რ-გნ-დ-ი-თ** სიმინც. „ჩვენ მიწით დავფარავდით სიმინდს“.

დო-ფო-ბ-რ-გნ-დ-ი-თ

PRV-R1-S1-R1-TH-IMPF-PM-PL

შდრ. საწყისი **ფორ-უა** “ფარვა”

4. კლიტიკები

სახელს და ზმნას დართული აქვს კავშირ-კლიტიკები (-ნი/-ნ; -და), **-ნი** ქართული „რომ“ კავშირის ეკვივალენტად შეიძლება ჩავთვალოთ, **-და** კი „თუ“-ს ტოლფარდად. ნიშანდობლივია ის, რომ კავშირ-ენკლიტიკები შეთვისებულია სახელთან ან ზმნასთან და დამოუკიდებლად არ დგას.

(14) ართ დღამა შეილებ არმა ტბა **ძირე-ნი**. „ერთ დღეში შეგიძლია, ყველა ტბა რომ ნახო“.

(15) შეილებუ სამყაროშ დასაწყისშა დუგორთინუას თი **ადგილეფქ-ნი**. “შესაძლოა, სამყაროს დასაწყისში რომ დაგაბრუნოს იმ ადგილმა“.

(16) ქუგოკო ათე აყვავილება, იფრელ **ქობირენ-და**, ივლისიმ სოდგარენ ვითიმე, სექტემბერიმ ვითმა ილნე. “თუ გინდა ეს აყვავილება ნახო, სადღაც სექტემბრის ათამდე შესაძლებელია მისვლა“.

(17) სკუა ქოქუნდ **თისგ-და**, ეთი კოს უჩაში სკუა ქოქუნდა, თიქ ოკო გამნიხვამუკონი. “შვილი თუ ჰყავდა მას, იმ კაცს უფროსი შვილი თუ ჰყავდა, მაგას უნდა ელოცა“.

დასკვნა

დასკვნის სახით შეიძლება ითქვას, რომ Flex-ი (Field Works Language Explorer) არის პროგრამა, რომელიც ენას სხვადასხვა კუთხით შეისწავლის, დამუშავებისა და სამომავლოდ მისი დაცვის საშუალებას იძლევა. ამ პროგრამით სიტყვის ფონეტიკური ანალიზი, ფორმების მორფემებად დაშლა, გლოსირება, კონკრეტული ენის ლექსიკონისა თუ გრამატიკული წესების შემუშავება, ენის დიგიტალიზაციაა შესაძლებელი. ამ პროცესში კი ახალი ფორმებისა თუ აქამდე უცნობი გრამატიკული მარკერების აღმოჩენაც შეიძლება.

პროგრამამ საშუალება მოგვცა უფრო თვალსაჩინო გამხდარიყო ის განსხვავებები, რომლებიც არსებობს, ერთი მხრივ, ზუგდიდურ-სამურზაყანულსა და მეორე მხრივ, სენაკურ-მარტვილურ ფორმებს შორის. მაგალითად, პირველისგან განსახვავებით (ზუგდიდურ-სამურზაყანული), მეორეში (სენაკურ-მარტვილური) შეიძლება შეგვხვდეს გრძელი ი სახელობითი ბრუნვის ნიშნად (-იი - თარგამეულიი), მიცემითის ნიშნად გრძელი -შე (-შეე - სამარგაჲოშეე), ან მრავლობითი რიცხვი შეიძლება გამოიხატოს გრძელი ხმოვნითი ნაწილით,

-ეფ (ბაღანეფ - ბავშვები) მარკერით. I სუბიექტური პირის ნიშნების შემთხვევაშიც ცხადად შეინიშნება სხვაობა მეგრულის ამ ორ დიალექტურ ფორმას შორის.

ლიტერატურა:

1. გერსამია, რ. (2009). იოტიზაციის საკითხისთვის მეგრულში. ენათმეცნიერების საკითხები. თბილისი. თბილისის უნივერსიტეტის გამომცემლობა.
2. გერსამია, რ. ლობჟანიძე, ი. სხულუხია, თ. წულაია, ნ. (2023). მეგრული ტექსტების დოკუმენტირება და ანოტირება FLEx -ში. ენა და კულტურა, 29:30-38.
3. Gersamia, R. (2022). The Morphonemics of Verbal Prefixes in Megrelian. Paris 8: [Presses Universitaires de Vincennes](#) PUV: 37-59.
4. გუდავა, ტ. გამყრელიძე, თ. (1981). თანხმოვანთკომპლექსები მეგრულში, აკაკი შანიძეს. თბილისი.
5. კარტოზია, გ. გერსამია, რ. ლომია, მ. ცხადაია, თ. (2011). მეგრულის ლინგვისტური ანალიზი. თბილისი. მერიდიანი.
6. Kobalava I., Gersamia R. (2020). Space and Motion in Language Representation (Analyses of Megrelian and Laz Linguistic Data), Tbilisi. 2020, Ilia State University. 2020.
7. ქობალავა, ი. (2020). სივრცისა და მოძრაობის ენობრივი რეპრეზენტაცია (მეგრული ენის მონაცემთა ანალიზი). თბილისი. ილიას სახელმწიფო უნივერსიტეტი.

Using Flex (Field Works Language Explorer) program for Megrelian texts

Tsulaia Nino
Skhulukhia Tamuna
Ilia State University, Tbilisi

Abstract

Modern technologies are actively used in the field of humanities. It became essential to apply them in linguistics as well. In the wake of technological development, it became possible to create language databases, which have been contributing to the representation and preservation of languages in a diachronic and synchronic way. It became topical to create language corpora, which makes it possible to preserve the language in the same form as it functions at a given moment. Language corpora are especially timely for unwritten, endangered languages found all over the world and also in Georgia. One of them is Megrelian, an unwritten language, which can be said to be at the edge of extinction. Our future goal is related to creating a Megrelian language corpus, which will include earlier and contemporary fieldwork data gathered from various districts of Samegrelo. The creation of the corpus will make it possible to study and investigate Megrelian broadly; to explore its sociocultural and linguistic (phonological, morphological) tendencies.

With the passage of time, the grammatical issues of languages are changing. Thus, it is important to capture and solve them. It is appropriate to use the corpus for the lexicographic purposes as well. Within the framework of the project, together with the compiling of the corpus, our goal is to create an electronic and printed dictionary. Defined tasks are performed and the fieldwork data is analyzed using the program FLEx (Fieldworks Language Explorer). The main stages of the entire working process are: obtaining and recording the aimed fieldwork data; describing the data and providing phonological annotations; compiling the data in the program and providing morphological analysis. The corpus will contribute to the protection of the language

under analysis. This corpus will be of significant help to linguists and researchers.

Keywords: Megrelian texts; morphological annotation in FLEx; Megrelian language corpus.