

მეგრული ტექსტების დოკუმენტირება და ანოტირება FLEx-ში¹

რუსუდან გერსამია, ირინა ლობჯანიძე,

თამუნა სხულუხია, ნინო წულაია

ილიას სახელმწიფო უნივერსიტეტი

rgersamia@iliauni.edu.ge, irina_lobzhanidze@iliauni.edu.ge tamuna.skhulukhia.2@iliauni.edu.ge,

nino.tsulaia.1@iliauni.edu.ge

<https://doi.org/10.52340/lac.2023.05>

მეგრული უმწერლობო ქართველური ენაა და ფუნქციონირებს მისი ორი დიალექტის, ზუგდიდურ-სამურზაყანოულისა და სენაკურ-მარტვილურის, სახით.

სტატიაში მეგრული ენის თანამედროვე ზეპირ მონაცემები, დამუშავებულია და გაანალიზებულია მეგრულენოვანი ჩანაწერების ლექსიკური და გრამატიკული მონაცემები FLEx-ის (FieldWorks Language Explorer) პროგრამის გამოყენებით. მოცემული ეტაპისათვის FLEx-ის ბაზაში შეტანილი და გაანალიზებულია 49 ტექსტი (ტოკენი - 38519, უნიკალური ფორმა -26610, წინადადება -3480). დიალოგები და მონოლოგები ჩაწერილია ორივე დიალექტურ არეალში და თემატურად მრავალფეროვანია. ყველა ტექსტს ახლავს თავისუფალი თარგმანი ქართულ და ინგლისურ ენებზე.

მორფოსინტაქსური ანალიზი მოიცავს შემდეგ სტრიქონებს:

1. *Word base* – წინადადებების დაყოფა სიტყვებად ავტომატურია.
2. *Word ipa* სტრიქონი - მეგრული ტექსტის საერთაშორისო ფონეტიკური ანბანური ჩანაწერის მომზადება ონლაინ კონვერტორში, რომელშიც შემუშავებულია კონვერტაციის სამი ძირითადი მიმართულება (გერსამია, ლობჯანიძე 2022). პირველი გულისხმობს ქართულ დამწერლობაზე დაფუძნებული მეგრულიდან საერთაშორისო ფონეტიკურ ანბანზე (IPA) გადასვლას, მეორე – მეგრულიდან ISO 9984 რომანიზაციის სქემაზე და მესამე გამყრელიძე-მაჭავარიანის ტრანსლიტერაციის სისტემაზე (1965 წ.) და პირუკუ.
3. *Morphemes* - სიტყვების მორფოლოგიური სეგმენტირება მანუალური პროცესია. სტრიქონში ფორმათა (ძირი, ფუძე, აფიქსი) მორფოსინტაქსური მონაცემი მანუალურად ხდება. სტრიქონი ასახავს ტექსტში ფიქსირებულ მონაცემებს.

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

4. მორფოსინტაქსური ანოტირება სამსტრიქონიანი და შეიცავს: *Lex. Entries, Lex. Gloss, Lex. Gram. Information*-ს და სემიმაწულური პროცესია. ინფორმაცია მორფმათა ვარიანტების, ამოსავალი მორფემების შესახებ აღწერილია *Lex. Entries*-ში, ხოლო *Lex. Gloss.*-ში მოცემულია მორფოსინტაქსური ანოტაცია. ტეგები ემყარება ლაიფცივის გლოსირებასა და ევროტიპის სპეციფიკაციას. ახალი ტეგები მუშავდება მეგრული ენის მორფოსინტაქსის გათვალისწინებით.

ამავე სტრიქონში ითარგმნება ძირული მორფემები ინგლისურად. სტრიქონი *Gram. Information* უზრუნველყოფს სემენტირებულ მონაცემებთან დაკავშირებული გრამატიკული კატეგორიების შესახებ ინფორმაციის ასახვას.

5. POS - მეტყველების ნაწილთა მიხედვით სიტყვების ანოტირების ველია.

ველზე მოპოვებული მეგრული ტექსტების მორფოსინტაქსური ანოტირება *FLEX*-ში საფუძვლად დაედება მეგრული ენის ფორმალურ გრამატიკას და ონლაინ ლექსიკონს.

შესავალი

მეგრული ქართველური ჯგუფის უმწერლობო ენაა. იგი ზეპირი სახით ფუნქციონირებს ორ, ზუგდიდურ-სამურზაყანოულ და სენაკურ-მარტვილურ, დიალექტად სამეგრელო-ზემო სვანეთის რეგიონში. ენგურსგადმელ, ამჟამად - დევნილ, ქართველთა მეგრული გაფანტულია საქართველოს სხვადასხვა რეგიონში. სამურზაყანოული ნარჩუნდება ოკუპირებული აფხაზეთის ტერიტორიაზეც გალში დარჩენილ მოსახლეობაში;

ენათა თანამედროვე ლინგვისტური, სოციოლინგვისტური, კულტურული აღწერა და კვლევა ეფუძნება მის დოკუმენტირებას - ვიდეოსა და აუდიოს დახმარებით მდგრადი ფაილების შექმნას, ჩანაწერების ანოტაციების მომზადებას მათი ხანგრძლივი გამოყენების მიზნით. „ენათა დოკუმენტირების“ შედეგად მიღებული „დოკუმენტირებული ენა“² თავისი არსით გულისხმობს საერთაშორისო ენობრივ და ტექნოლოგიურ სტანდარტებზე დაყრდნობით დამუშავებულ ენას, რომელიც ინტერნეტ სივრცეში განთავსებით ხდება საყოველთაოდ ღია და წვდომადი რესურსი მკვლევართათვის.

² გასული საუკუნის 90-იანი წლებიდან გამოჩნდა ენათმეცნიერების ახალი ქვედარგი, რომელსაც უწოდეს "ენის დოკუმენტაცია" ან "დოკუმენტური ლინგვისტიკა" (Himmelman 1998, Lehmann 2001, Austin 2010), რომელიც მიმართული იყო საფრთხეში ანუ გადაშენების პირას მყოფი ენებისკენ. ტერმინების გაგება განსხვავდებოდა ადრინდელი ისტორიული გაგებისგან, რომელიც მოიცავდა გრამატიკებისა და ლექსიკონების, ტექსტური კრებულების შექმნას იმ პარამეტრების მქონე ტექნიკის გამოყენებით, რომლებიც ვალიდური იქნება მონაცემთა შემდგომი ტექნოლოგიური დამუშავებისათვის (Austin 2013).

ველზე მონაცემთა შეგროვების შემდგომ მათი ციფრული დამუშავებისათვის აუცილებელ პირობას წარმოადგენს საველე ჩანაწერების გადატანა, შენახვა-დაცვა და მართვა შესაბამისი კომპიუტერული უზრუნველყოფის გამოყენებით. აღნიშნული მიზნების მისაღწევად გამოიყენება მრავალკომპონენტური პროგრამული რესურსი - FLEx (FieldWorks Language Explorer)³. ამ კომპონენტთა ურთიერთზემოქმედებით ხდება მონაცემების მართვა.

FLEx (FieldWorks Language Explorer) მონაცემთა მართვის პროგრამაა, მონაცემების ველზე მოპოვებიდან ონლაინ გრამატიკის, ლექსიკონისა და მონაცემის ბაზის შექმნამდე. მონაცემთა მართვის არსი ისაა, რომ აღნიშნული პროგრამა ეხმარება მკვლევარს ჩანაწერების ორგანიზებასა და მათი შემდგომი ლინგვისტური ანალიზის დროს.

მეგრული ენის მონაცემთა დამუშავება FLEx-ში

პროექტის მიზნებიდან და ამოცანებიდან გამომდინარე მეგრული ენის თანამედროვე ზეპირი მასალის ჩანაწერების აღწერა, დამუშავება და ანალიზი ლექსიკურ და გრამატიკულ (resp. მორფოსინტაქსურ) დონეზე ხორციელდება FLEx-ის (FieldWorks Language Explorer) პროგრამის გამოყენებით. საველე ჩანაწერები, რომლებიც სხვადასხვა ასაკისა და სქესის ინფორმანტებისაგან 2022 წლის ზაფხულში მოპოვებულ იქნა მარტვილის, ჩხოროწყუსა და წალენჯიხის სოფლებში, მოიცავს ჯამურად 22 საათიან ზეპირ დიალოგებსა და მონოლოგებს, რომლებითაც იფარება ორივე დიალექტური არეალი. ჩანაწერები თემატურად მრავალფეროვანია, შინაარსობრივად მეგრულთა რწმენა-წარმოდგენებს, ყოფა-ცხოვრებას, ტრადიციებსა და მათთან დაკავშირებულ რიტუალებს; შელოცვებს, დამოკიდებულებებს ენასა და კულტურასთან; ტოპონიმებს, ხალხურ ეტიმოლოგიებსა და სათავგადასავლო ამბებს უკავშირდება.

მოცემული ეტაპისათვის FLEx-ის ბაზაში შეტანილი და გაანალიზებულია 49 ტექსტი (ტოკენი - 38519, უნიკალური ფორმა - 26610, წინადადება - 3480). ყველა ტექსტს ახლავს თავისუფალი თარგმანი ქართულ და ინგლისურ ენებზე.

FLEx-ის ბაზაში ტექსტების შეტანის შემდეგ პროგრამა ავტომატურად ანიჭებს ნომრებს აბზაცებს (1, 2, 3...) და აბზაცში წინადადებებს (1.1, 1.2, 1.3...). წინადადების გამყოფ ნიშნად აღიქმება წერტილი. ვინაიდან მეგრული ზეპირი ენაა, წინადადების სიგრძის დადგენა პრობლემურია და მთლიანად დამოკიდებულია ინტონაციაზე. წერტილიდან წერტილამდე ტექსტის ინფორმაციული სტრუქტურა საკმაოდ რთული და მოცულობითია; გახლეჩილი ჰიპოტაქსური სინტაქსური კონსტრუქციის შიგნით დამოკიდებული წინადადებების სახით

³ იხ. <https://software.sil.org/fieldworks/>, ბოლო წვდომის თარიღი 10.05.2023

დეტალურად არის ჩაშლილი მთავარი წინადადების წევრებით გადმოცემულ აზრს მიბმული დეტალური ინფორმაცია. გაბმულ თხრობას ახასიათებს რამდენიმე სპეციფიკური ნიშანი, რომელიც აზრის გახლეჩის და წერტილის ფუნქციას ასრულებს: ა) პაუზის *ოდო, ონი* “ზოდა” ნაწილაკები, რომლებიც, როგორც წესი, იწყებს წინადადებას; ბ) *დო* “და” კავშირი, რომელიც წინადადებების გადაბმისათვის გამოიყენება, როცა არ არის პარატაქსულად მაკონსტრუირებელი ან ერთგვარ წევრთა დაჯგუფების ფუნქციით გამოყენებული. ამიტომ მოცულობითი წინადადებების სეგმენტირებისათვის აღნიშნული ნიშნებიც გამოიყენება.

FLEX-ში ანალიზის ნაწილი შემდეგ სტრიქონებს მოიცავს:

ცხრილი 1.

5.2 Word base	ათე		თამადაქ		მოჭყუდუმი		
Word ipa	atɛ		tamadak		moɟʰqʷduɟi		
Morphemes	ათე	-A0	თამადა	-ქ	მოჭყუდუ	-ში	-ი
Lex. Entries	ათე	-ი ₁	თამადა	-ქ ₁	მოჭყუდუ	-იმ	-ი ₂
Lex. Gloss	this.PROX	NOM	toastmaster	ERG	bride	GEN	EMPH
Lex. Gram. Info.	dprn	pro-form:(Case)	cn	n:(Case)	n	n:(Case)	n:(Emphatic vowel)
Word Cat.	dprn		cn		cn		

1. **Word base – წინადადებების დაყოფა სიტყვებად ავტომატურია.**

FLEX-ის პროგრამა ასევე ავტომატურად შლის წინადადებებს სიტყვებად და მონაცემთა სტრიქონულად დამუშავების საშუალებას იძლევა. მეგრული ტექსტის სიტყვებად დაყოფისას (პარსირებისას) მნიშვნელოვანია გათვალისწინებული იყოს კლიტიკების საკითხი სახელებსა და ზმნებში. სპეციალურ ლიტერატურაში გაზიარებულია აზრი ქართველურ ენათამცნირებაში ზმნურ სტრუქტურაში შემავალი პრეფიქსული ენკლიტიკური ნაწილაკები *ვა-*, *ქო-* და ზმნასა თუ სახელურებს დართული ენკლიტიკური კავშირებისა (*-ნი*, *-და*) და სხვათა სიტყვის, კითხვითობისა და ჩაკითხვის ნაწილაკების, *რენ* “არის” მეშველი ზმნის სხვადასხვა დროის ფორმის ნაცვეთი ნაწილის / ეროზიული ვარიანტის (*-ე*, *-დ*, *-ი/-იი*) შესახებ, თუმცა მეგრული ენის თანამედროვე აუდიო ჩანაწერები ერთმარცვლიანი ელემენტების ერთი მახვილის ქვეშ გაერთიანების ტენდენციას კვლავ ავლენს; კერძოდ, თავს იჩენს მეტყველების ნაწილებით, ერთმარცვლიანი კითხვითი ნაცვასახელების ზმნასთან კომბინირებით მიღებული სტრუქტურები: *მემიჩქუდა* < **მა მიჩქუ* და “მე როგორც ვფიქრობ”, *მუმიჩქ* < მუ მიჩქ “რა ვიცი”, *მამზე* < *მუამზე* < *მუ ამზე* “რა ამბავი”, *მიირდ* < *მი რდ* “ვინ იყო”... აღნიშნული უკავშირდება უფრო პროსოდისა და არა გრამატიკალიზაციის პროცესს.

FLEX-ის პროგრამაში კლიტიკები ბაზაში შესატან ტექსტში სწორად უნდა აისახოს, რომ პროგრამამ შეძლოს მათი შემდგომი სწორი დამუშავება. კლიტიკისათვის პროგრამის მიერ

შემოთავაზებული ტოლობის ნიშნის [=] გამოყენება კი სვამს საკითხს ყველა ცალსახად არაფლექსიური და არადერივაციული აფიქსის ინტერპრეტაციის შესახებ.

2. **Word ipa სტრიქონი - მეგრული ტექსტის საერთაშორისო ფონეტიკური ანბანური ჩანაწერის** მომზადება ონლაინ კონვერტორში, რომელშიც შემუშავებულია კონვერტაციის სამი ძირითადი მიმართულება (გერსამია, ლობჯანიძე 2022). პირველი გულისხმობს ქართულ დამწერლობაზე დაფუძნებული მეგრულიდან საერთაშორისო ფონეტიკურ ანბანზე (IPA) გადასვლას, მეორე – მეგრულიდან ISO 9984 რომანიზაციის სქემაზე და მესამე გამყრელიძე-მაჭავარიანის ტრანსლიტერაციის სისტემაზე (1965 წ.) და პირუკუ. მუშაობის პროცესში, ძირითადად, პირველი გამოიყენება და, შესაბამისად, ახალი ქართულის ბგერათსისტემას მეგრულისათვის სპეციფიკური რამდენიმე ბგერა ემატება, კერძოდ, ფარინგალური /ყ/, ნეიტრალური /ჟ/ ხმოვანი, იოტა /ა/ და უბრჯგუ /უ/.

3. ოთხსტრიქონიანი **მორფოსინტაქსური ანოტირება** შემდეგ სტრიქონებს მოიცავს:

3.1. Morphemes - მორფემები

3.2. Lex. Entries - ლექსიკური ჩანაწერები

3.3. Lex. Gloss - ლექსიკური გლოსები

3.4. Lex. Gram Info - ლექსიკურ-გრამატიკული ხასიათის ინფორმაცია (იხ. ცხრილი 1)

3.1. Morphemes - სიტყვების მორფოლოგიური სეგმენტირება არავტომატური პროცესია და გულისხმობს ბმული ფლექსიური მორფემების დაშლას, რის შემდეგაც მიიღება ძირეული და აფიქსური (პრეფიქსი, სუფიქსი, კონფიქსი, ინფიქსი) მორფემები. სეგმენტირება ემყარება ზოგად მორფოლოგიურ ცოდნას, ენათმეცნიერების თეორიებს, ქართველურ ენებში აფიქსთა გამოყოფის, სიტყვის სტრუქტურირების ზოგად პრინციპებსა და მეგრული ენის მორფოლოგიის წესებს (ჩიქობავა 1936; კარტოზია და სხვ. 2011; ქირია და სხვ. 2015).

საკითხები, რომლებიც FLEx-ში ანალიზის დაწყებამდე ითხოვს გადაწყვეტას და ცალსახა პასუხებს:

ა. ფლექსიური და დერივაციული აფიქსების გამიჯვნა. ამ ეტაპისათვის FLEx-ის ანალიზის მიღმა რჩება ლექსიკური დერივაცია, რომელიც ახალი სალექსიკონო ერთეულების / ფუძეების წარმოშობას უკავშირდება, მაგრამ ზოგადად ფლექსიისა და დერივაციის საკითხების გამიჯვნა მნიშვნელოვანია პროგრამაში ანალიზისათვის; მაგალითად, ისეთ საკითხებზე წინასწარ შეთანხმება, აფიქსთა რომელ ჯგუფს განეკუთვნება თანდებულები სახელებში, ან ზმნისწინები ზმნებსა და სახელზმნებში? შდრ. ი. ქობალავას თვალსაზრისი „ზმნისწინები მეგრულში უფრო სადერივაციო ფორმანტებია“ (ქობალავა, 2020:13).

მეგრულში, ასევე, ნაკლებად დგას ფლექსიური (პოლიფუნქციური) მორფემების საკითხი, რადგან იგი მკვეთრად აგლუტინაციური ენაა თუნდაც ყველა დანარჩენ ქართველურ ენასთან შედარებით, რაც იმას ნიშნავს, რომ ერთი აფიქსი მხოლოდ ერთი ფუნქციის გამომხატველია (მელიქიშვილი, 2010:91; Kobalava, Gersamia, 2020:12), სამაგიეროდ აქტუალურია მორფემათა ფუზიურობა.

ბ. ნულოვანი ჩანაწერები - [^0]. სამი სხვადასხვა ფუნქციით არის შესაძლებელი მათი გამოჩენა სტრიქონში: 1. როცა ნული რომელიმე მორფემის ალომორფია; 2. როცა მორფემა გამოტოვებულია; 3. როცა ტექსტში დასტურდება „კვეცილი სტილი“ (გუდავა 1975:356-360).

გ. კლიტიკები. სახელებსა და ზმნებს ნაწილაკური წარმოშობის ენკლიტიკები და პროკლიტიკები, ასევე კავშირ-ენკლიტიკები ერთვის. ახალი ტექსტები ერთმარცვლიანი კითხვითი ნაცვალსახელების პროკლიტიკად ქცევის და დამხმარე ზმნების ენკლიტიკებად ქცევის სისტემურ მაგალითებს ავლენს.

დ. ფუნქციურად გარდაქმნილი ფორმები. გრამატიკული მორფემები ზოგჯერ სიტყვებს ფუნქციურად ცვლიან, მაგ.: ზედსართავი სახელის ვითარებითი ბრუნვის ფორმა ვითარების ზმნიზედაა თუ ზედსართავი სახელი. შესაბამისად, იბადება კითხვა უნდა თუ არა გამოიყოს ვითარებითი ბრუნვის ნიშანი? ან აქვს თუ არა ბრუნვები ზმნიზედებს მეგრულში და შესაძლებელია თუ არა მათი გამოყოფა პარადიგმატულად? დაიშალოს თუ არა ნაცვალსახელები ან ზმნიზედები ფორმოზრივად გამჭვირვალე შემადგენელ კომპონენტებად, რომლებიც ფუნქციური თვალსაზრისით მთლიან ერთეულს წარმოადგენენ, მაგ.: ნაცვალსახელები: *თიშგურშენი* „იმიტომ“, *თეშმაგვარი* „ამგვარი“, *თეჯგურა* „ასეთი“, *მითინი* „არავინ“, *ნამუთი* „რომელიც“, *მიდგარენი* „ვიღაც“, *ნამდგარენი* „რომელიღაც“ და სხვ. ზმნიზედები: *სოდგარენი* „სადღაც“, *სოლედგაინი* „საიდანღაც“ და სხვ. როგორ გაფორმდეს მსაზღვრელი სიტყვა, მიენიჭოს თუ არა ბრუნვა, როცა საზღვრულთან პრეპოზიციამი კვეცილად არის წარმოდგენილი? და ა.შ.

3.2. **Lex. Entries** - ნახევრად ავტომატური სტრიქონია და იძლევა ინფორმაციას აფიქსურ და ძირეულ მორფემათა ვარიანტების (დიალექტური, ფონეტიკური, არტიკულაციური), ალომორფთა ამოსავალი მორფემების (3) შესახებ, ამავე სტრიქონში შესაძლებელია გამოტოვებული მორფემების აღდგენა, მათ შორის ნულოვანი ძირებისა, რომლებიც [^0] ჩანაწერის სახით გაფორმდა წინა სტრიქონში (იხ. ცხრილი 1); ამავე სტრიქონში შესაძლებელია აღდგეს ძირი თავდაპირველი, ამოსავალი მიმდევრობით, მაგ. (1) (2):

(1) მე- მ- კუნ- „მიგვყავს“
PRV O1 ცუნ+sp. var. of

- (2) ცაფ / ააფ „თააში“
ლაყაფი+dial. var. of
- (3) გიფ- ნო- ვ -ე „დაყაყდა თურაე“
გილა--+dial. var. of EV ცვ EV

აღნიშნული ველი, ერთი მხრივ, მნიშვნელოვანია ამოსავალი მორფოსტრუქტურის, მეორე მხრივ კი, ზედაპირულ სტრუქტურაში რეპრეზენტირებული მორფემების ჩვენებისთვის, რომლითაც მეგრულში შეიძლება განისაზღვროს საკმარისი და აუცილებელი მორფემები ამა თუ იმ მოდელისათვის, რაც, თავის მხრივ, ღირებულია ენის ეკონომიურობის პრინციპების დასადგენად ენაში.

3.3. Lex. Gloss და 3.4. Lex. Gram. Information - ავტომატური სტრიქონებია, რომლებიც ლექსიკურ და გრამატიკულ ინფორმაციას იძლევა თითოეული დაშლილი სეგმენტის შესახებ. მორფოსინტაქსური ანოტაცია ემყარება ლაიფციგის გლოსირების წესებსა და ევროტიპის ინსტრუქციებს ტეგების სია შემუშავდა მეგრული ენის მორფოსინტაქსის გათვალისწინებით. მუშაობის პროცესში მოხდა მათი წინასწარი დამუშავება და FLEx Grammar-ის ნაწილში შეტანა. ანოტირებისას ფანჯარა არსებული მონაცემებიდან არჩევის, ახალი მნიშვნელობების, ვარიანტების შექმნის საშუალებას იძლევა ან ნაცნობ სტრუქტურებს თავად აანალიზებს ავტომატურად.

ამავე სტრიქონში ინგლისურად ითარგმნება ძირეული მორფემები. თარგმანი არის ნეიტრალური და მოკლე. სტრიქონი Gram. Information უზრუნველყოფს სეგმენტირებულ მონაცემებთან დაკავშირებული გრამატიკული კატეგორიების შესახებ ინფორმაციის ასახვას.

4. ბოლო სტრიქონი POS (Part of speech annotation) **მეტყველების ნაწილთა** მიხედვით სიტყვების **ანოტირების ველია**, რომელიც ინფორმაციას გვაწვდის არა დაშლილი ფორმის, არამედ მთლიანი სიტყვის შესახებ. შესაძლებელია მეტყველების ნაწილთა ათივე ჯგუფის კონკრეტულ ტიპებად ჩაშლა და ამ სახით FLEx-ის ანალიზში გამოტანა, მაგ.: ნაცვალსახელი, რიცხვითი სახელები, ზმნიზედები, ნაწილაკები და ა.შ. წარმოდგენილია მათი შინაარსობრივი ჯგუფების მიხედვით.

FLEx-ში შეტანილ მონაცემთა რედაქტირება შესაძლებელია სხვადასხვა გარემოდან, მაგ. სამ პარალელურ ფანჯარაში ლესიკონის, ტექსტის და მორფოლოგიურად ანოტირებული ტექსტის ჩვენებით, ასევე ცალკეული რედაქტორებიდან როგორც ლექსიკონის, ასევე გრამატიკის ნაწილში.

დასკვნა

FLEx-ის პროგრამის მოთხოვნები მკაფიოდ მიჯნავს ენის სინქრონიულ და დიაქრონიულ მოცემულობებს, ცალსახად ორიენტირებულია დესკრიფციულ მორფოლოგიაზე და მისი

საანალიზო ერთეული სიტყვაა. საბოლოო ჯამში, ფლექსის პროგრამა საშუალებას იძლევა ენის მორფოლოგია დავინახოთ სისტემაში და რაოდენობრივად ბევრი მონაცემის მორფოლოგიური ანალიზით (სეგმენტური მორფოლოგიური ანალიზით) სისტემაში დავინახოთ მეგრული ენის მორფოლოგია, გადავამოწმოთ დამკვიდრებული თვალსაზრისები, ვნახოთ ახალი ტენდენციები.

საბოლოოდ, სტრიქონებში მოცემული ლექსიკური და მორფოლოგიური ინფორმაცია პროგრამიდან ლექსიკონისა და ფორმალური გრამატიკის გენერირებას ემსახურება.

ლიტერატურა:

1. გუდავა ტ., - ქართული ხალხური სიტყვიერება, მეგრული ტექსტები, I, პოეზია, გამომცემლობა „თბილისის უნივერსიტეტი“. თბილისი 1975.
2. კარტოზია გ., გერსამია რ., ლომია მ., ცხადაია თ., - მეგრულის ლინგვისტური ანალიზი. გამომცემლობა “მერიდიანი”. თბილისი 2011.
3. ლობჯანიძე ი., - ქართული ენის მორფოსინტაქსური ანოტირებისა და სასრული პოზიციის მორფოლოგიური ანალიზის პრინციპები. ილიას სახელმწიფო უნივერსიტეტის გამომცემლობა, თბილისი 2021
4. მახარობლიძე თ., - დერივანტების შესახებ, ლინგვისტური წერილები IV, გამომცემლობა “წიგნის ფაბრიკა”. თბილისი 2018.
5. მელიქიშვილი ი., - ფლექსიურობის (ფუზიურობის) შესახებ ქართველურ ენებში, ენათმეცნიერების საკითხები, I-II. 2010.
6. ქირია ჭ., ეზუგბაია ლ., მემიშიში მ., ჩუხუა მ., - ლაზურ-მეგრულის მორფოლოგია I. გამომცემლობა “მერიდიანი”. თბილისი 2015
7. ქობალავა ი., - სივრცისა და მოძრაობის ენობრივი რეპრეზენტაცია (მეგრული ენის მონაცემთა ანალიზი). ილიას სახელმწიფო უნივერსიტეტის გამომცემლობა. თბილისი 2020.
8. ჩიქობავა არნ., - ჭანურის გრამატიკული ანალიზი. სსრკ მეცნიერებათა აკადემიის საქართველოს ფილიალის გამოცემა. ტფილისი. 1936
9. Austin, P. K. - Current issues in language documentation. In Peter K. Austin (ed.) Language Documentation and Description, Volume 7:12-33. London: SOAS. 2010.
10. Austin, P. K. - Language documentation and meta documentation. Keeping Languages Alive (Documentation, Pedagogy and Revitalization). Eds. M.C Jones and S. Ogilvie. Cambridge University Press.2013
11. Gamkrelidze, T., Machavariani, G. - The system of sonants and ablaut in the Kartvelian languages. Tbilisi: Science. 1965.
12. Gersamia, R., Lobzhanidze, I. - Mingrelian Converter.

<https://irinalobzhanidze.com/megrelian/converter/converter.html>. Accessed 15 June, 2022.

13. Himmelmann, N. P. - Documentary and descriptive linguistics. *Linguistics* 36,161-195. 1998.

14. Kobalava I., Gersamia R. - *Space and Motion in Language Representation (Analyses of Megrelian and Laz Linguistic Data)*, Tbilisi. 2020, Ilia State University. 2020.

15. Lehmann, Ch. *Language documentation. A program*. In Walter Bisang (ed.) *Aspects of Typology and Universals*, 83-97. Berlin: Akademie Verlag. 2001.

Rusudan Gersamia, Irina Lobjanidze, Tamuna Skhulukhia, Nino Tsulaia

Ilia State University

Documentation and annotations of Megrelian texts in FLEx

Abstract

Megrelian is an unwritten Kartvelian language subdivided into two dialects Zugdidi-Samurzaqano and Senaki-Martvili. Following the approaches widely used in language documentation, the Megrelian texts are described and analyzed by means of FieldWorks Language Explorer (FLEx).

The FLEx provides interlinear glossing of data. At this moment the Megrelian database includes 49 analyzed texts (38510 tokens, 26610 unique forms, 3480 sentences). Dialogues and monologues are recorded in both dialect areas and include information about the beliefs, traditions, rituals etc. All texts are represented with free translations in Georgian and English. The analysis includes the following:

1. Word_{base} - division into words done automatically;
2. Word_{ipa} - representation of Megrelian text in IPA done by the online converter (Gersamia, Lobzhanidze 2022) developed to convert Megrelian based on Georgian script into the International Phonetic Alphabet (IPA), the ISO 9984 romanization scheme, the Gamkrelidze-Machavariani transliteration system (1965) and vice versa;
3. Morphemes - morphological segmentation of words is done manually. The word is subdivided into root and affixes representing fixed linguistic data;
4. The information on morphemes is given in Lex. Entries, the morphological glossing - in Lex. Gloss, while Lex. Gram. Info describes PoS-es and their appropriate categories. Glossing is done following the Leipzig Glossing Rules and Eurotyp Guidelines;
5. Word_{cat} - information on PoS-es is given separately as well.

The morphosyntactic annotation of Megrelian texts collected during the fieldwork and annotated by means of FLEx will be used as a base for the grammar and online dictionary.

რეცენზენტი: აკადემიკოსი ვაჟა შენგელია