

მარინა ბერიძე, დავით ნადარაია

დიდი ქართული დიალექტური ლექსიკოგრაფიული ბაზა და ქართული დიალექტების ელექტრონული ატლასი (GEDA)

შესავალი: პროექტი „დიდი ქართული დიალექტოლოგიური ლექსიკოგრაფიული ბაზა და ქართული დიალექტების ელექტრონული ატლასი (GEDA)¹“ არის გაგრძელება მრავალწლიანი სამუშაოს, რომელსაც ზოგად სახელად „საქართველოს ლინგვისტური პორტრეტი“ ვუწოდებთ და რომლის ფარგლებშიც შეიქმნა ორი კორპუსული რესურსი: ქართული დიალექტური კორპუსი და ქართული სამეცნიერო მეტაენის კორპუსი“.

ქართული დიალექტების ლექსიკოგრაფიული ბაზა ქართული დიალექტური კორპუსის ცალკე კომპონენტად შეიქმნა. ის არის ტექნოლოგიური სისტემა, კორპუსის პლატფორმაში ჩაშენებული პროგრამული უზრუნველყოფა, რომლის საშუალებითაც შესაძლებელია ლექსიკოგრაფიული მასალის მობილიზება, მისი სტრუქტურირება, ლინგვისტური, მეტალინგვისტური, ენციკლოპედიური და სხვ. ხასიათის ინფორმაციით აღჭურვა, ქართული დიალექტების კრებსითი ლექსიკონისა და მასზე დაფუძნებული სხვადასხვა ტიპის ლექსიკონების შექმნა.

მეორე სიახლე, რომელიც შემოიტანა პროექტმა ქართულ სამეცნიერო სივრცეში, ეს არის კორპუსული მონაცემების კარტოგრაფიული ვიზუალიზაციის შესაძლებლობა.

ენობრივი მონაცემების კარტოგრაფირების კონცეფცია – ისტორიული და თანამედროვე რაკურსები: ქართული დიალექტოლოგიისა და დიალექტოგრაფიის განვითარების სტრატეგია გასული საუკუნის დასაწყისიდან მუშავდებოდა. ჯერ კიდევ 1911წ. პეტერბურგის ქართველ სტუდენტთა სამეცნიერო წრემ ი. ჯავახიშვილის ხელმძღვანელობით შეადგინა პირველი ინსტრუქცია დიალექტური მასალების დოკუმენტირებისათვის, 1927 წელს ქართულმა საენათმეცნიერო საზოგადოებამ შეიმუშავა ანალოგიური პროგრამა-ინსტრუქცია

¹ კვლევა განხორციელდა შოთა რუსთაველის საქართველოს ეროვნული სამეცნიერო ფონდის მხარდაჭერით (პროექტი „დიდი ქართული დიალექტოლოგიური ლექსიკოგრაფიული ბაზა და ქართული დიალექტების ელექტრონული ატლასი (GEDA)“ – № 217008).

[ჩიქობავა 2008: 34]. გასული საუკუნის 60-იან წლებში დიალექტოლოგიისა და დიალექტოგრაფიის საკითხებს სპეციალური საკოორდინაციო თათაბირი მიემდგვნა [1960]. დარგის განვითარების ერთგვარი პროგრამა იყო 1963 წელს დაბეჭდილი ვარლამ თოფურიას სტატია „ქართველურ ენათა შესწავლის მდგომარეობა და ამოცანები“. ამ საპროგრამო ნაშრომში ავტორი ცალკე გამოყოფს ქართული დიალექტოგრაფიის ახალ ამოცანას – დიალექტურ მოვლენათა აღსაწერად „კარტოგრაფიული მეთოდის მომარჯვებას“ [თოფურია 1963: 178]. სტატიიდან ჩანს, რომ იყო მცდელობა კარტოგრაფირებისთვის მასალის მოპოვების ცენტრალიზებული და კოორდინირებული სისტემის შესაქმნელად, რომელიც წარუმატებელი აღმოჩნდა. ქართულ სამეცნიერო საზოგადოებას რაიონებში დაუგზავნია გიორგი ახვლედიანის მიერ შედგენილი ანკეტა, ასევე, ახალი ქართული ენის კათედრა პედაგოგიკის ინსტიტუტის მეშვეობით ცდილობდა ანკეტური მასალების მოპოვებას, მაგრამ „ამას დიდი შედეგი არ მოჰყოლია“. მკვლევარი იმედოვნებდა, „იქნებ კორესპონდენტთა ქსელის შექმნამ გააცხოველოს საქმე“, თუმცა ესეც არ მოხდა და 80-იან წლებამდე საქმე წინ აღარ წასულა. 1984 წელს გამოიკა კითხვარი „ქართული დიალექტური ატლასის მასალისათვის“ [კიზირია და სხვ. 1984]. კიდევ უფრო გვიან, 90-იან წლებში, არნ. ჩიქობავას სახელობის ენათმეცნიერების ინსტიტუტში შეიქმნა „ლინგვისტური ატლასის ლაბორატორია“, რომლის მიზანი იყო საქართველოს ყველა რეგიონში ზემოხსენებული კითხვარით მასალის მოპოვება და მიღებული შედეგების კარტოგრაფირება. პროექტი „საქართველოს ლინგვისტური პორტრეტი“ ამ ლაბორატორიის ნამემკვიდრევი პრობლემატიკაზე ამოიზარდა.

ლინგვისტური ატლასების – ეროვნული ენობრივი რუკების შექმნა უკავშირდება ლინგვისტური გეოგრაფიის წარმოშობას და განვითარებას და მე-19 საუკუნის შუა წლებიდან იღებს სათავეს. თეორიულ საფუძვლად ლინგვისტურ გეოგრაფიას ენობრივი უწყვეტობისა და ტალღათა თეორიის კონცეფციები დაედო. ის იკვლევდა და კარტოგრაფიულად ასახავდა ცალკეულ ენობრივ მოვლენებს, შეისწავლიდა ამ მოვლენათა გეოგრაფიულ განფენილობას – იზოგლოსებს. ეს კვლევები ემყარებოდა მოსაზრებებს, რომ ენობრივი მოვლენები

უწყვეტი ტალღების სახით ვრცელდებიან ცენტრიდან პერიფერიისაკენ. ლინგვისტური გეოგრაფია შეიქმნა ამ თეორიების კარტოგრაფიული ვიზუალიზაციის უზრუნველსაყოფად, მას ჰქონდა აგრეთვე ამბიცია, დაედგინა დიალექტებს შორის საზღვრები.

ლინგვისტური გეოგრაფიის დისციპლინურ ჩარჩოში შეიქმნა ევროპულ ენათა უმდიდრესი ლინგვისტური ატლასები. მათი აგების პრინციპი სხვაობდა კარტოგრაფირების მიზნისა და ენობრივი ლანდშაფტის გათვალისწინებით, თუმცა კონცეფცია ერთი იყო – წინასწარ შემუშავებული კითხვარების მიხედვით შევსებული ანკეტების მონაცემები გადადიოდა რუკებზე, ერთნაირი ან მსგავსი მოვლენების გამაერთიანებელი გეოგრაფიული ხაზები – იზოგლოსები ასახავდნენ დიალექტური მოვლენების იზოლაციის ან საერთოობის ხარისხს. მიუხედავად სხვადასხვა თეორიული საყრდენისა, რუკების პრაქტიკული გამოყენება, ფაქტობრივად, ვერ გასცდა დიალექტთა შორის საზღვრების დადგენის საკითხს: არსებობს თუ არა საერთოდ დიალექტური საზღვარი, თუ არსებობს, რამდენად მკვეთრი შეიძლება ის იყოს, რა ხდება ენობრივი სივრცის იმ მონაკვეთზე, რომელსაც „დიალექტური გარდამავალი ზონები“ უწოდეს?!

ამ კითხვებზე პასუხი უნდა გაეცა იმ ატლასსაც, რომლის შექმნის იდეაც ჯერ კიდევ 60-იანი წლებიდან გაჩნდა ჩვენს სამეცნიერო სივრცეში და რომელზეც ზემოთ ვისაუბრეთ. როცა ჩვენმა სამუშაო ჯგუფმა გადაწყვიტა ლინგვისტური ატლასის მასალების ბაზაზე შეექმნა „საქართველოს ლინგვისტური პორტრეტი“ და ამისთვის გამოეყენებინა თანამედროვე ტექნოლოგიები, სრულიად აშკარა გახდა, რომ სერიოზულად უნდა გადასინჯულიყო მეთოდოლოგია და პროგრამა. რამდენად აქტუალური იყო ამ დროისთვის მე-19 საუკუნის ფრანგული და გერმანული დიალექტოლოგიის წიაღში წარმოშობილი და იმ კონკრეტულ დროსა და გარემოს მორგებული „არაპირდაპირი ანკეტირების“ მეთოდით მასალის მოპოვება და მისი პირდაპირ გადატანა რუკაზე? რომელ კითხვებზე უნდა გაეცა პასუხი ასეთ რუკებს? რომელი ენობრივი მოვლენების კარტოგრაფირება იყო უფრო მნიშვნელოვანი და საშური?

ვარლამ თოფურია თავის წერილში აღნიშნავდა: „ქართველურ ენათა მასალის მიხედვით ირკვევა, რომ დიალექტები ისეთ მკვეთრ ფონეტიკურსა და გრამატიკულ ნაირსახეობებს არ იძლევა, რომ მათი კარტაზე გადატანით ატლასი შეიქმნასო“ [178] და კარტოგრაფირებისთვის შესაფერისად უფრო ლექსიკურ თავისებურებებს მიიჩნევდა. თუმცა არც მის ნაშრომში და არც მოგვიანებით დიალექტური კითხვარის შემქნელთა მიერ არ ყოფილა ჩამოყალიბებული ის ძირითადი კითხვები, რომელზეც ქართულ დიალექტურ ატლასს უნდა გაეცა პასუხი.

დიალექტების საზღვრის დადგენა?

გარდამავალი დიალექტური ზონების გამოვლენა?

დიალექტური (ენობრივი) უწყვეტობისა და „ტალღათა თეორიის“ კონცეფციათა შემოწმება?

ქართული დიალექტები, რომლებიც ეთნიკურ-ტერიტორიული და ლინგვისტური პრინციპით არის გამოყოფილი, რასაკვირველია, ავლენენ მეზობელ დიალექტებთან ე.წ. „გარდამავალ ზონებს“, მაგრამ ეს ზონები საკმაოდ კარგად არის აღწერილი ლიტერატურაში და არც დიალექტთა შორის საზღვრების დადგენა გამხდარა სადაო ქართულ დიალექტოლოგიაში. თუმცა მე-20 საუკუნის ინტენსიური მიგრაციული პროცესების გამო [ბერიძე 2011: 7-8] იმდენად შეიცვალა ქართული დიალექტების რეალური განფენილობა, რომ თითზე ჩამოსათვლელი დიალექტებიღა დარჩა თავის ისტორიულ საზღვრებში. კითხვარი, რომლითაც იკრიბებოდა მასალა ყველა ქართულ დიალექტში, ფაქტობრივად ასახავდა არა რეალურ განფენილობას, არამედ ადასტურებდა დიალექტურ ფორმას დიალექტის ისტორიულ ფარგლებში. ასეთი იყო ინსტრუქცია მასალის შეკრების დროს.

ენობრივ მოვლენათა კარტოგრაფირების ჩვენი კონცეფცია, ზემოთგანხილულ მიზეზთა გამო ბრმად არ დაყრდნობია გეოლინგვისტიკის ტრადიციულ მეთოდოლოგიას.

„ლინგვისტური პორტრეტის“ სამუშაო პრიორიტეტები ისეთი თანმიმდევრობით დალაგდა, რომ ჯერ შექმნილიყო ქართული დიალექტური სივრცის რეპრეზენტაციული კორპუსული მოდელი და შემდგომ მომხდარიყო ამ

მოდელის ცოდნის ინსტრუმენტად ქცევა სახვადასხვა ტექნოლოგიური კომპონენტების, მათ შორის კარტოგრაფიული ვიზუალიზაციის საშუალებების დამატებით. გეგმა ასეთი იყო:

1. ქართული დიალექტური სივრცის ტექსტური პორტრეტის შექმნა;
2. ქართული დიალექტური სივრცის ლექსიკოგრაფიული პორტრეტის შექმნა;
3. მე-20 საუკუნის ქართული მიგრაციული მარშრუტების აღწერა (მიგრაციული პორტრეტის შექმნა)
4. არსებული მონაცემების კარტოგრაფიული ვიზუალიზაცია.

აი, ამ ინფორმაციულ და ტექნოლოგიურ ბაზაზე დაყრდნობით შეიძლება უკვე შეიქმნას სხვადასხვა ტიპის სტატისტიკური თუ დინამიკური, ინტერაქტიური რუკები, რომლებიც შეიძლება იყოს გენერირებული სხვადასხვა ლინგვოგეოგრაფიული კონცეპტის თუ ამოცანის შესაბამისად – იქნება ეს ვენკერისეული 42 კითხვის ("Wenkersätze") კარტოგრაფირება (DIVA, 2001-2009), ენობრივ სტრუქტურათა გეოგრაფიული გავრცელების (WALS) ან გაცილებით ახალგაზრდა მიმართულების – დიალექტომეტრიის შედეგების რუკაზე ვიზუალიზაცია და სხვ.

ამოცანა „მიგრაციული პორტრეტის შექმნა“ ნაწილობრივ განხორციელდა ტექსტური ბაზის მეტამახასიათებლების მონაცემებში მიგრაციული ბლოკის დამატებით. გარდა ამისა, პროექტის ერთ-ერთ ეტაპზე ჩვენ შევქმენით „მიგრაციების ბაზა“, რომელიც სულ მალე იქცა მიგრაციული და ლინგვისტური პროცესების კარტოგრაფირების მთავარ ინსტრუმენტად. 2007 წელს გამოქვეყნდა ამ ბაზის გამოყენებით შექმნილი რამდენიმე რუკა: მიგრაციები კახეთში, მიგრაციები ცხენისწყლის ხეობაში, სამცხე-ჯავახეთის იმერული მიგრაციები, მიგრაციები სვანეთიდან, აჭარიდან და სხვ. [ბერიძე, უგულავა, 2007].

ამჟამად კორპუსულ და მიგრაციულ ბაზებზე დაყრდნობით შესაძლებელია შემდეგი სახის ინფორმაციის კარტოგრაფიული ვიზუალიზაცია:

- ლემის კარტოგრაფიული ვიზუალიზაცია

- სიტყვაფორმის კარტოგრაფიული ვიზუალიზაცია
- ლემასთან დაკავშირებული ყველა სიტყვაფორმის კარტოგრაფიული ვიზუალიზაცია
- მეტატექსტური მონაცემების კარტოგრაფიული ვიზუალიზაცია
- მიგრაციული მონაცემების კარტოგრაფიული ვიზუალიზაცია

კარტოგრაფირებისთვის გამოყენებულ იქნა უახლეს გეოინფორმაციულ სისტემებზე დაყრდნობილი ინსტრუმენტები, რის შედეგადაც შეიქმნა არა ტრადიციული სტატიკური ვიზუალი ამა თუ იმ კორპუსული მონაცემის კარტოგრაფირებისათვის, არამედ, დინამიკური, ინტერაქტიური რუკები, რომლებიც გენერირდება მომხმარებლის მიერ საძიებო პარამეტრების საჭირო კონფიგურირებით. სანიმუშო მოდელად ამ სამუშაოებისთვის შესწავლილი და გათვალისწინებული იყო ცნობილი კარტოგრაფიული პროექტები: WALS (<http://wals.info/>) და DIWA (<http://www.diwa.info/> და სხვ.), თუმცა ჩვენ არ შეგვიქმნია არცერთი ამ სისტემის ანალოგი. მათგან განსხვავებით ჩვენს „ატლასში“ კარტოგრაფიული ინტერესი ფოკუსირებული არ არის კონკრეტულ ენობრივ მოვლენებზე. კარტოგრაფიული ვიზუალიზაციის სისტემა დაყრდნობილია კორპუსის მეტაანოტირებისა და მიგრაციული ბაზების გეოგრაფიულ კომპონენტზე და უახლესი გეოინფორმაციული სისტემების ამ ბაზებთან დაკავშირებით იქმნება მთელი კორპუსული მონაცემებისა და არა ცალკეული ენობრივი ფაქტების ინტერაქტიური, დინამიკური რუკები.

ქართული დიალექტური ლექსიკოგრაფია და ქდკ-ს ლექსიკოგრაფიული კონცეფცია.

დიდი ლექსიკოგრაფიული ბაზის მოსამზადებლად სპეციალურად შემუშავდა უნივერსალური, ინტერაქტიური, ლექსიკოგრაფიული სისტემა – მოქნილი ინსტრუმენტი კორპუსზე დაყრდნობილი ახალი დიალექტური ლექსიკონების შესაქმნელად. სისტემა წარმოადგენს კორპუსის არქიტექტურაშივე ჩაშენებული ლექსიკონების რედაქტორს – კომპილატორს – ერთგვარ „ლექსიკოგრაფიულ ლაბორატორიას“. ლექსიკოგრაფიული ბაზა ქდკ-ს

პლატფორმის დამოუკიდებელი ნაწილია. ის არის ლექსიკოგრაფიული მასალის შეკრების, დამუშავების და ლექსიკონებად გარდაქმნის მოქნილი ინსტრუმენტი, რომელიც თავს უყრის ლექსიკას ყველა დიალექტური ლექსიკონიდან და ტექსტური ბაზიდან და თითოეული ლექსემის ყველა კონტექსტური რეალიზაციის, აგრეთვე სიხშირული მონაცემის გათვალისწინების საშუალებას იძლევა სალექსიკონო სტატიის შექმნის პროცესში.

ლექსიკოგრაფიული ბაზა მოიცავს: გამოქვეყნებული ბეჭდური ლექსიკონების ელექტრონულ ვერსიებს; სხვადასხვა დიალექტურ ექსპედიციაში მოპოვებულ ლექსიკოგრაფიულ ინფორმაციას; სხვადასხვა ავტორის მიერ გამოქვეყნებულ ან ხელნაწერის სახით არსებულ ლექსიკოგრაფიულ მასალას; ლინგვისტური, ეთნოგრაფიული და სხვ. ხასიათის ნაშრომებიდან ამოკრებილ ლექსიკოგრაფიულ მასალას; კორპუსის ტექსტებიდან ამოკრებილ მასალას.

ლექსიკოგრაფიული ბაზა, რომელიც სხვადასხვა ლექსიკოგრაფიული პრინციპითა და ტრადიციით შექმნილ ლექსიკონებს აერთიანებს, სამუშაო მასალაა ქდკ-ს დიდი დიალექტური ონლაინლექსიკონის შესაქმნელად.

დღეისათვის ქართული დიალექტების ლექსიკოგრაფიური დოკუმენტირების დონე არათანაბარია. გვაქვს რამდენიმე დიალექტის ვრცელი, აკადემიური ლექსიკონი: ინგილოური, ქართლური, იმერული, ხევსურული, თუშური, აჭარული..., ზოგი დიალექტისთვის რამდენიმე, სხვადასხვა მოცულობისა და შედგენილობის, სხვადასხვა პრინციპით შედგენილი ლექსიკონი არსებობს (გურული, იმერული, ინგილოური...) ზოგი დიალექტი კი არ არის სათანადო ლექსიკოგრაფიული პროდუქციით უზრუნველყოფილი და მხოლოდ ცალკეული პუბლიკაციების, მცირე მოცულობის სალექსიკონო მასალების ან სულაც ხელნაწერი მასალების სახით გვაქვს მათზე წარმოდგენა (ლექხუმური, ფშაური, მთიულური, ჯავახური.... არსებობს აგრეთვე ქართული დიალექტების კრებსითი ლექსიკონიც [დლონტი: 1984], რომელიც აერთიანებს გამოცემული დიალექტური ლექსიკონების მასალას.

ამ პროდუქციას არ ახასიათებს ლექსიკოგრაფიული პრინციპების ერთიანობა, განსხვავებულია სალექსიკონო ერთეულის წარმოდგენის წესი,

სალექსიკონო სტატიის სტრუქტურა, ლინგვისტური და ენციკლოპედიური ინფორმაციის გადმოცემის პრინციპი და სხვ. ქართულ დიალექტურ ლექსიკოგრაფიაში დამკვიდრებული ტრადიციით დიალექტური ლექსიკონები ასახავენ მხოლოდ მოცემული დიალექტისთვის სპეციფიკურ, სალიტერატურო ენისაგან განსხვავებულ ლექსიკას. სალიტერატურო ენასთან საერთო ლექსიკა დიალექტურ ლექსიკონში ხვდება მხოლოდ მაშინ, როცა ის სრულად ან ნაწილობრივ განსხვავებული მნიშვნელობით დასტურდებოდა დიალექტში.

ქართული დიალექტური ლექსიკონებისთვის ეს დიალექტური „პურიზმი“ იმდენად მნიშვნელოვან მოთხოვნად იქცა, რომ ზოგი ავტორი ერიდებოდა არა მხოლოდ სალიტერატურო ენასთან, არამედ სხვა დიალექტებთან საერთო ლექსიკური ერთეულების შეტანას ლექსიკონებში.

ლექსიკის რა ნაწილი აისახა დღემდე გამოცემულ ქართულ დიალექტურ ლექსიკონებში? ეს არის:

1. დიალექტური სიტყვები, რომლებიც არ / აღარ გამოიყენება სალიტერატურო ენაში
2. ფორმით ან მნიშვნელობით განსხვავებული სიტყვები, შესიტყვებები, ფრაზეოლოგიზმები.
3. სალიტერატურო ენის ის სიტყვები, რომელთაც განსხვავებული მნიშვნელობა აქვს დიალექტებში (ან სალიტერატუროსთან საერთო მნიშვნელობასთან ერთად განსხვავებული მნიშვნელობაც აქვთ).
4. არქაიზმები, რომლებიც შემოინახა დიალექტმა.
5. დიალექტების საერთო ლექსიკა (რამდენიმე სხვადასხვა დიალექტში დადასტურებული, ერთი და იმავე ან განსხვავებული მნიშვნელობით).
6. უცხო ენიდან შემოსული ლექსიკა
7. დიალექტებისთვის დამახასიათებელი მორფემები და მოდალური ელემენტები.
8. ზოგიერთი ტიპის საკუთარი სახელი [მარტიროსოვი:139-148.]

როგორც წესი, ქართულ დიალექტურ ლექსიკონებში არ (ან ძალიან იშვიათად) ხვდებოდა: საკუთარი სახელები(მეტსახელები, გვარები, ტოპონიმები; რეგულარული ფონეტიკური ნაირსახეობები, უცხოურიდან შემოსული ისეთი სიტყვები, რომლის დიალექტური ფორმაც დასტურდება („დიალექტური ბარბარიზმები“)...

დღემდე გამოცემულ ლექსიკონებში არაერთგვაროვანია სალექსიკონო სტატიის სტრუქტურირების პრინციპებიც: არ არის ერთგვაროვნება არც მაკრო და არც მიკროსტრუქტურების წარმოდგენისას – განსხვავებულია სალექსიკონო ერთეულის წარმოდგენისა და განმარტების ბლოკში ლექსიკოგრაფიული ინფორმაციის სტრუქტურირების პრინციპებიც.

ზოგიერთი ლექსიკონი სალექსიკონო ერთეულად წარმოადგენს რამდენიმე ფონეტიკურ ვარიანტს, ან ასევე სალექსიკონო ერთეულის სტრუქტურულ ველში იძლევა სხვადასხვა გრამატიკულ ვარიაციას (მაგალითად, სალექსიკონო ერთეულად წარმოდგენილია ზმნის მესამე პირის ფორმა, რომელსაც მოსდევს მასდარისა და სხვადასხვა დროის მიმდევობის ფორმები, სახელებთან მითითებულია ზოგი ირიბი ბრუნვის ან მრავლობითის ფორმა და ა. შ.).

ასეთივე ნაირგვარობაა განმარტების ინფორმაციული ველების სტრუქტურირების თვალსაზრისით: განსხვავებულია ლინგვისტიკური, ენციკლოპედიური, ილუსტრაციული, მიმართებითი (რეფერენციული) და სხვა ინფორმაციის მოწოდების საშუალებები.

ამ მრავალფეროვნების დაძლევის და ერთიან პრინციპებზე დაფუძნებული დიდი დიალექტური ლექსიკონის შექმნის იდეა რეალური გახადა კორპუსული და კომპიუტერული ტექნოლოგიების განვითარებამ. ამ იდეის ნაწილი იყო ქართული დიალექტური კორპუსის შექმნაც და მის ბაზაზე შემუშავებული ლექსიკოგრაფიული კონცეფციაც.

ქდკ-ს ლექსიკოგრაფიული კონცეფცია ითვალისწინებს:

- არსებული ლექსიკონების ინფორმაციის მეორეულ გადამუშავებას, უნიფიცირებას და სტრუქტურირებას;

- ახალი ლექსკური ერთეულების დამატებას კორპუსიდან, ახალი საველე მასალებიდან და სხვა ლექსიკონებიდან;
- ლექსიკონის გამდიდრებას საილუსტრაციო, ლინგვისტური და ენციკლოპედიური ინფორმაციით.

ეს სამუშაო ერთმანეთის პარალელურად, ეტაპობრივად და ერთიანი გეგმით სრულდება. კონცეფცია გულისხმობს დიდი დიალექტური ბაზის ფორმირებას, რომელიც ქართულ დიალექტურ ლექსიკას წარმოადგენს როგორც ერთიანი – ინტეგრირებული სახით, ისე ცალკეული დიალექტების ონლაინლექსიკონების სახით.

ქდკ-ს ლექსიკოგრაფიული რედაქტორი წარმოადგენს ინსტრუმენტს (ხელსაწყოს) რომელიც ლექსიკოგრაფიული ბაზის ინფორმაციის „შენახვისა“ და მისი სტრუქტურირების საშუალებას იძლევა.

ლექსიკოგრაფიული ინფორმაცია რედაქტორში შესაძლებელია როგორც პირდაპირი დამატების გზით, ისე წინასწარ მომზადებული სტრუქტურირებული ცხრილების ატვირთვის სახით (Excel). ლექსიკოგრაფიული რედაქტორი არ ითხოვს მომზადებული ფაილების რაიმე სხვა ფორმატს და შესაბამისად, მასზე მომუშავე ლექსიკოგრაფების ან ოპერატორების განსაკუთრებულ სპეციალურ კვალიფიკაციას.

ლექსიკოგრაფიული რედაქტორი მოიცავს სხვადასხვა სარედაქციო ველს და ფუნქციონალს: 1) შენახვის ველი_ ასეთი ველი რამდენიმეა და მათში საწყისი ლექსიკოგრაფიული წყაროს ინფორმაცია ინახება. ამ ინფორმაციაზე დაყრდნობით ხდება ახალი სალექსიკონო სტატიის ფორმირება, რედაქტირება, შემოწმება და ა. შ. იმის გამო, რომ ხშირად ერთი სტატია სხვადასხვა ლექსიკოგრაფიულ წყაროს ეყრდნობა, ლექსიკოგრაფიული რედაქტორის „საცავში“ რამდენიმე ველი გვაქვს თითოეული ასეთი წყაროს შესანახად; 2)ლემის დამატების ველი; 3)შესატყვისის დამატების ველი; 4)განმარტების დამატების ველი; 5)ენციკლოპედიური ინფორმაციის დამატების ველი; 6)სიტყვის უცხოური წარმომავლობის შესახებ ინფორმაციის დამატების ფუნქციონალი (რომელი ენიდანაა ნასესხები); 7)სიტყვის უცხოური დაწერილობის დამატების ველი (როგორ იწერება სიტყვა იმ ენაზე,

რომლიდანაც არის ნასესხები; 8)ინფორმაცია სიტყვის სტრუქტურის შესახებ (ენკლიტიკა, კომპოზიცია, მრავალკომპონენტური სიტყვა, რთული გრამატიკული ფორმა, მარტივი შესიტყვება); 9)ინფორმაცია ლექსიკური ტიპის შესახებ (ნეიტრალური, ტროპული, დარგობრივი); 10)ინფორმაცია სემანტიკური კლასის შესახებ; 11)ინფორმაცია ლექსიკოგრაფიული წყაროს / წყაროების შესახებ; 12)ლემის გრამატიკული მახასიათებელი; 12)ლემის გრამატიკული (ფორმაწარმოებითი) ვარიაცია; 13)ლემის სიტყვაწარმოებითი ვარიაცია; 14)ლემის ფონეტიკური ვარიაცია; 15)მითითება სინონიმზე; 16)მითითება სხვა ლექსიკურ ერთეულზე; 17) ილუსტრაცია; 18)ილუსტრაციის თარგმანი ან კომენტარი; 19)ილუსტრაციის წყარო და სხვ.

ქდკ-ს ლექსიკოგრაფიული რედაქტორი იძლევა სალექსიკონო სტატიის შექმნის, რედაქტირების, ცვლილების, გამოქვეყნების შესაძლებლობას. მარკირებულია სალექსიკონო სტატიის სარედაქციო სამუშაოთა სხვადასხვა საფეხურიც: დამუშავების, რედაქტირებისა და გამოქვეყნების საფეხური.

დამუშავების საფეხურზე ლექსიკოგრაფიული რედაქტორის ველებიდან გააქტიურებულია მხოლოდ ორი – **ლემა და შენახვის ველი**. ეს იმას ნიშნავს, რომ რედაქტორში შეტანილია ლემა (რომელიც დიალექტური კორპუსის საერთო სიტყვანში აისახება) და აგრეთვე „გროვდება“ ინფორმაცია ყველა შესაძლო ლექსიკოგრაფიული წყაროდან.

რედაქტირების საფეხურზე იწყება არსებული ლექსიკოგრაფიული ინფორმაციის სტრუქტურირება და განთავსება შესაბამის ველებში.

გამოქვეყნების საფეხურზე შექმნილი სრული სალექსიკონო სტატია ხელმისაწვდომი ხდება მომხმარებლისთვის.

დღეისათვის ლექსიკოგრაფიულ ბაზაში 10-ზე მეტი დიდი და მცირე ლექსიკონია გაერთიანებული (60000-მდე სიტყვა), ამათგან ექვსი გამოქვეყნებული და ხელმისაწვდომია (ფერეიდნული, ინგილოური, იმერული, ლაზური, „ჩვენებურების“, თუშური), დანარჩენები კი დამუშავების სხვადასხვა ეტაპზეა, თუმცა საერთო ძიება მათში უკვე შესაძლებელია როგორც სიტყვის ან მისი

ფრაგმენტის, ასევე გრამატიკული, სემანტიკური, სტრუქტურული და სხვა მახასიათებლებით [ბერიძე 2017:107-122].

მრავალდონიანი ანოტირების სისტემა ლექსიკოგრაფიულ ბაზაში ენობრივი და მეტამახასიათებლების სამომხმარებლო კონფიგურირების საშუალებას იძლევა. მოქნილი საძიებო ინსტრუმენტის საშუალებით შესაძლებელია სხვადასხვა დიალექტის სხვადასხვა ტიპის ლექსიკონების გენერირება (მაგალითად: ნასესხები სიტყვების ლექსიკონი, კომპოზიტების ლექსიკონი, ამა თუ იმ სემანტიკური თუ გრამატიკული ჯგუფის სიტყვების ლექსიკონი და სხვ.).

„დიდი ქართული დიალექტური ლექსიკოგრაფიული ბაზის“ იდეა იყო მცდელობა, ერთ სივრცეში წარმოედგინა კორპუსული და ლექსიკოგრაფიული კომპონენტი, როგორც კვლევის რესურსი. ამასთან, მოეხდინა კორპუსული მონაცემების კარტოგრაფიული ვიზუალიზაცია.

დღეს ეს ბაზა უკვე არის დიალექტების ლექსიკოგრაფირების ეფექტური ინსტრუმენტი, რომლის განვითარება ორი მხრივ უნდა წარიმართოს. ესაა მუდმივი მასალობრივი შევსება და რაც შეიძლება მეტი თანამედროვე ტიპის ლექსიკოგრაფიული პროდუქციის შექმნა.

ბიბლიოგრაფია:

1. ბერიძე მარინა, ქართული ეთნოსოციალური სინამდვილის აქტუალიზაციის საკითხისათვის, Folia Caucasica, ფრანკფურტი მაინზე, თბილისი, 2011. 1-33გვ.
2. თოფურია ვარლამ, ქართველურ ენათა დიალექტების შესწავლის მდგომარეობა და ამოცანები, ქართული ენის სტრუქტურის საკითხები, თბილისი, 1963.
3. კიზირია, ი. ქავთარაძე, მ. ჭაბაშვილი, თ. უთურგაიძე (შემდგენლები) კითხვარი ქართული დიალექტური ატლასის მასალისათვის, თბილისი, 1984.
4. მარინა ბერიძე, ირაკლი უგულავა, ელექტრონული რუკები (15.11.2018) <http://titus.uni-frankfurt.de/ssgg/ssgg04.HTM#dProjekt>

5. მარტიროსოვი არამ, ქართული დიალექტური ლექსიკის შესწავლისა და ლექსიკონების შედგენის ძირითადი საკითხები, იკე, ტ. XXIII, 1985 139-148.
6. მარტიროსოვი,
7. ქართული დიალექტოლოგიის საკითხები, საკოორდინაციო თათბირის მასალები, თბილისი, 1960.
8. ქართული დიალექტური კორპუსი (15.11.2018) <http://corpora.co/>
9. ქართული ეროვნული კორპუსი <http://gnc.gov.ge>
10. ქართული სამეცნიერო მეტაენის კორპუსი (15.11.2018) <http://sml.corpora.co/>
11. ღლონტი ალექსანდრე, ქართულ კილო-თქმათა სიტყვის კონა, თბილისი, 1984.
12. ჩიქობავა არნოლდ, ზოგადი ენათმეცნიერების შესავალი, თბილისი, 2008
13. ჯორბენაძე ბესარიონ, ქართული დიალექტოლოგია, I, 1989.
14. ჯორბენაძე, ქართული დიალექტოლოგია 1, თბილისი, 1989.
15. Beridze Marina, David Nadaraia, Lexicographic Potential of the Georgian Dialect Corpus. In *Proceedings of the XVII EURALEX International Congress, Lexicography and Linguistic Diversity*. Pp. 300-309, 2016.
16. Beridze Marina, David Nadaraia, The Corpus of Georgian Dialects, In *Proceedings of the NLP, Corpus Linguistics, Corpus Based Grammar Research. Fifth International Conference Smolenice, Slovakia. Tribun*. Pp. 25-35, (2009)
17. Beridze Marina, lia Bakuradze, Zakharia Pourtskhvanidze, David Nadaraia, Georgian Dialect Corpus: Linguistic and Encyclopedic Information in Online Dictionaries. In *Journal of Linguistics/Jazykovedný casopis. The Journal of Ludovít Štúr Institute of Linguistics, SAV. N.68/2*. Pp.109-121, 2017.
18. Beridze Marina, Liana Lordkiphanidze, David Nadaraia, Dialect Dictionaries in the Georgian Dialect Corpus, *Logic, Language, and Computation/ XIV Springer*. Pp. 82–96, 2015.
19. Beridze, Liana Lordkiphanidze, David Nadaraia, The Georgian Dialect Corpus: Problems and prospects. In *Proceedings of the conference on Historical Corpora Challenges and Perspectives, Frankfurt*. Pp. 323–333, 2015
20. Digitaler Wenker-Atlas _ DIWA (<http://www.diwa.info/>)

21. Szmrecsanyi Benedikt Corpus-based dialectometry: aggregate morphosyntactic variability in British English dialects', In *International Journal of Humanities and Arts Computing 2 (1-2)*. Pp. 279–96, 2008.
22. The World Atlas of Language Structures _ WAL <http://wals.info>
23. Trudgill Peter, 'Linguistic change and diffusion: description and explanation in sociolinguistic dialect geography', In *Language in Society 3 (2)*. Pp. 215–46, 1974.
24. Viereck, W., H. Ramisch, H. Händler, P. Hoffmann and W. Putschke (1991) *The Computer Developed Linguistic Atlas of England*. Tübingen: Niemeyer.

The Large Georgian Dialect Lexicographic Database and the Georgian Electronic Dialect Atlas (GEDA)

The project “**The Large Georgian Dialect Lexicographic Database and The Georgian Electronic Dialect Atlas GEDA**” is a continuation and elaboration of comprehensive activities undertaken within the framework of the project *Linguistic Portrait of Georgia*.

At its initial stages, significant platforms for representation of Georgian dialect linguistic data were designed: working and user applications of text corpora and electronic dictionaries, a two- million text base; the lexicographic base was initiated. The system of morpho-syntactic parsing was developed; partial annotation was conducted. Several electronic dictionaries were compiled and published.

Based on the dialect corpus and the lexicographic editor, the present project was aimed *at developing of the comprehensive Georgian dialect lexicographic base and at providing for the cartographic visualization of linguistic data*.

As an interactive instrument, the Georgian dialectological lexicographic base and its cartographic component (Electronic Atlas of Georgian Dialects) is a significant accomplishment both owing to great amount of information and its effective usage functions and to the fact that it is a open resource and accessible to everyone.

The base is apt to become a powerful teaching/learning and scholarly resource. It is applicable both at various levels of teaching/learning and as for scholarly purposes – as a tool for diverse linguistic and interdisciplinary studies.

Project activities were undertaken in two directions: 1)Developing of the comprehensive Georgian dailect lexicographic base and of the technological framework for the cartographic visualization; 2)Elaboration of the comprehensive Georgian dailect lexicographic base (publication of new electronic dictionaries, their multi-level annotation, compilation of electronic maps)

Elaboration of the markup system of dictionaries qualitatively changed opportunities of customary queries in them – a user is able to generate a desired dictionary from a sizeable massive of data and develop various sub-dictionaries (a dictionary of foreign words, a dictionary of compounds, a dictionary of proper names, a dictionary of idioms, etc.).

As a result of the implementation of the project, the information potential of the Georgian Dialect Corpus increased twice. A user can enjoy the universal and unique text and lexicographic reference system both with a wide and flexible query system and with unique opportunities to self-generate various kinds of dictionaries and to cartographically visualize corpus data.