

# მიმოხილვა

## REVIEW ARTICLES

მარინა ბერიძე

*თსუ არნოლდ ჩიქობავას სახელობის  
ენათმეცნიერების ინსტიტუტი*

დავით ნადარაია

*დამოუკიდებელი მკვლევარი, დოქტორი*

### ქართული დიალექტური კორპუსის ლექსიკოგრაფიული კონცეფცია

#### შესავალი

ქართული დიალექტური კორპუსი (ქდკ) არის საინფორმაციო-საცნობარო პლატფორმა, რომელიც, შეიქმნა პროექტის „საქართველოს ლინგვისტური პორტრეტი“ ფარგლებში და წარმოადგენს ქართული დიალექტების კორპუსული დოკუმენტირების პირველ ცდას.

ქდკ რამდენიმე კომპონენტისგან შედგება. ესენია: საკუთრივ კორპუსი, ტექსტების ბიბლიოთეკა, ქდკ-ის ლექსიკოგრაფიული ბაზა /დიალექტური ონლაინლექსიკონები, მიგრაციების ბაზა. ამასთან, კორპუსში ინტეგრირებულია ტექსტური კორპუსისა და მიგრაციების ბაზის მონაცემების კარტოგრაფიული ვიზუალიზაციის ინტერაქციული სისტემა.

პროექტის განვითარების ბოლო საფეხურზე ქდკ-ის მთავარი პლატფორმის სტრუქტურაში ჩაშენებული ლექსიკოგრაფიული რედაქტორის ახალი ვერსიისა და კორპუსული მონაცემების დამუშავების სხვადასხვა ინსტრუმენტის განვითარების შედეგად, ყველა ეს კომპონენტი ჩამოყალიბდა ერთ მთლიან პროდუქტად, რომლის პირობითი სახელწოდებაცაა: **დიდი ქართული ლექსიკოგრაფიული ბაზა და ქართული დიალექტების ელექტრონული ატლასი GEDA** (ბერიძე და ნადარაია 2018).

მნიშვნელოვანი კორპუსული ტექნოლოგია, რომელიც ამ პროექტში დამუშავდა, არის უნივერსალური, ინტერაქციული, ლექსიკოგრაფიული სისტემა (პლატფორმა), რომელიც უკვე არის კარგი და მოქნილი ხელსაწყო კორპუსზე დაყრდნობილი ახალი დიალექტური ლექსიკონების შესაქმნელად. სისტემა წარმოადგენს კორპუსის არქიტექტურაში ჩაშენებული ლექსიკონების რედაქტორს – კომპილატორს – ერთგვარ „ლექსიკოგრაფიულ ლაბორატორიას“; რომელიც თავს უყრის ლექსიკას ყველა დიალექტური ლექსიკონიდან და ტექსტური ბაზიდან და თითოეული ლექსემის ყველა კონტექსტური რეალიზაციის გათვალისწინების საშუალებას იძლევა სალექსიკონო სტატიის შექმნის პროცესში.

პროექტზე მუშაობის დროს შესწავლილ და გათვალისწინებულ იქნა ლექსიკოგრაფიული მონაცემების მოდელირების რამდენიმე ტექნოლოგია (Lemon – The Lexicon Model for Ontologies (ლექსიკური მოდელი ონტოლოგიებისათვის); LexInfo – a declarative model for the lexicon-ontology interface (დეკლარაციული მოდელი ლექსიკონის ონტოლოგიების სამომხმარებლო ნიღბისათვის); SKOS – Simple Knowledge Organization System RDF Schema (ცოდნის მარტივი ორგანიზების სისტემა – RDF სქემა).

კიდევ ერთი ახალი ფუნქცია, რომელიც დაემატა პროდუქტს, არის კორპუსული მონაცემების კარტოგრაფიული ვიზუალიზაციის შესაძლებლობა.

კარტოგრაფირებისთვის გამოყენებულ იქნა უახლეს გეოინფორმაციულ სისტემებზე დაყრდნობილი ინსტრუმენტები, რის შედეგადაც შევძელით, შეგვექმნა არა ტრადიციული სტატიკური ვიზუალი ამა თუ იმ კორპუსული მონაცემის კარტოგრაფირებისათვის, არამედ დინამიკური, ინტერაქციული რუკები, რომლებიც გენერირდება მომხმარებლის მიერ საძიებო პარამეტრების საჭირო კონფიგურირებით. სანიმუშო მოდელად ამ სამუშაოებისთვის შესწავლილი და გათვალისწინებული გვექონდა ლინგვისტური მონაცემების კარტოგრაფირების ცნობილი სისტემები: WALS (<http://wals.info/>) და DIWA (<http://www.diwa.info/> და სხვ.), თუმცა არ შეგვიქმნია არცერთი ამ სისტემის ანალოგი, რაც კონცეპტუალური მოსაზრებით იყო განპირობებული. ჩვენი მიზანი იყო, კარტოგრაფიული ვიზუალიზაციის სისტემა დაგვეყრდნო კორპუსის მეტაანოტირების ბაზაზე და უახლესი გეოინფორმაციული სისტემების ამ ბაზებთან დაკავშირებით შეგვექმნა მთელი ბაზის და არა ცალკეული ენობრივი მოვლენების ინტერაქციული რუკები.

დღეისთვის ქართული დიალექტური კორპუსი შეიცავს: 2 054 705 სიტყვას და 477 129 სიტყვაფორმას. ამასთან, ცალკეული დიალექტების მასალა არაპროპორციულადაა წარმოდგენილი. ის დიალექტები,

რომელთა მასალის შევსება პარალელურად მიმდინარეობდა პროექტის სამუშაო ჯგუფის მიერ (ინგილოური, ფერეიდნული, თურქეთის ქართული დიალექტები, სამცხურ-ჯავახური, ქართლური, კახური, გურული, იმერული – შედარებით ვრცლად არის წარმოდგენილი, ხოლო დიალექტების ან ცალკეულ თქმათა ის ნაწილი, რომელთა მასალაც მხოლოდ „ქართული დიალექტოლოგიის“ მიხედვით გვაქვს დოკუმენტირებული, შესაბამისად, მწირად არის ასახული. ქდკ-ს ცალკეული დიალექტური ქვეკორპუსების მოცულობა ასეთია:

ქართული დიალექტები ( // თქმები): აჭარული – 24294, გურული – 26567, ებრაელების ქართული – 8840, თიანური – 6669, თუშური – 38496, იმერული – 38894, ინგილოური – 37398, კახური – 63253, ლეჩხუმური – 7799, მთიულურ-გუდამაყრული – 22708, მოხეური – 9866, რაჭული – 43562, სამცხურ-ჯავახური – 3814, ფერეიდნული – 50601, ფშაური – 23993, ქართლური – 42454, ყიზლარ-მოზდოკური – 1503, სამხრული კილოები (ჩვენებურების ქართული) – 53989, ხევსურული – 8070. სხვა ქართველური ენების დიალექტები: ლაზური – 10943.

### 1. ქდკ-ის ლექსიკოგრაფიული რედაქტორი

ქდკ-ის ლექსიკოგრაფიული რედაქტორი პლატფორმის დამოუკიდებელი ნაწილია. ის არის ლექსიკოგრაფიული მასალის შეკრების, დამუშავების და ლექსიკონებად გარდაქმნის ინსტრუმენტი.

ლექსიკოგრაფიული ბაზის ინფორმაციული წყაროებია: გამოქვეყნებული ბეჭდური ლექსიკონების ელექტრონულ ვერსიები; არასალექსიკონო სამეცნიერო რესურსებიდან ამოკრებილი ლექსიკოგრაფიული მასალები; ქართული დიალექტური კორპუსის ტექსტური მასივის ლექსიკა.

ქდკ-ის ლექსიკოგრაფიული ბაზის შევსება გამუდმებით მიმდინარეობს სხვა კორპუსული სამუშაოების პარალელურად, ამჟამად ლექსიკოგრაფიულ ბაზაში ქართული ენის სხვადასხვა დიალექტისა და ლაზურის 99 305 სალექსიკონო ერთეულია თავმოყრილი:

გურული – 2139, თუშური – 6893, იმერული – 11775, ინგილოური – 13990, კახური – 6741, ლაზური – 5948, ლეჩხუმური – 1688, მოხეური – 1255, სამცხურ-ჯავახური – 4480, ფერეიდნული – 5372, ქართლური – 5942, ჩვენებურების ქართული (თურქეთის ტერიტორიაზე გავრცელებული დიალექტები) – 9500, ხევსურული – 25508.

ამათგან სრული სახით გამოქვეყნებულია:

თუშური – 6815, იმერული – 7158, ინგილოური – 10423, ლაზური – 4374, ფერეიდნული – 5357, ჩვენებურების ქართული – 9490 და ლაზური – 594.

ლექსიკოგრაფიულ რედაქტორში პარალელურად მიმდინარეობს სხვადასხვა ტიპის სამუშაო ქდკ-ის დიდი დიალექტური ონლაინლექსიკონის შესაქმნელად.

როგორც ვნახეთ, დღეისათვის ქდკ-ის შვიდი ონლაინლექსიკონია გამოქვეყნებული. ეს ლექსიკონები შეადგენენ ერთიანი ქართველური ლექსიკონის ბაზის ნაწილს. ამ ბაზიდან მათი ცალკე ლექსიკონებად „დაგენირება“ საძიებო სისტემით ხდება.

ქდკ-ის ერთიანი ლექსიკოგრაფიული კონცეფციის მიხედვით მუშავდება და გამოსაქვეყნებლად მზადდება ახალი ონლაინლექსიკონები.

2018 წელს ფრანკფურტის წიგნის ბაზრობისთვის მომზადდა სამი კორპუსული ლექსიკონის ექსპერიმენტული გამოცემა, ესენია: თუშური, იმერული და ფერეიდნული დიალექტების კორპუსული ლექსიკონები (ბერიძე, ბაკურაძე და ნიკოლოზ ნახუცრიშვილი 2018; ბერიძე, შარაშენიძე და კიკნაძე 2018; ცოცანიძე, და ბერიძე 2018).

ექსპერიმენტული ლექსიკონების მომზადების მთლიანი ციკლი (გარდა საბოლოო კორექტურისა) მიმდინარეობდა ქდკ-ის ლექსიკოგრაფიულ რედაქტორში. ეს ციკლი სხვადასხვა ტიპის სამუშაოს მოიცავდა იმის მიხედვით, თუ რა იყო სალექსიკონო სტატიის ამოსავალი ლექსიკოგრაფიული წყარო: ბეჭდური ლექსიკონი, ლექსიკოგრაფიული კითხვარით მოძიებული / გადამოწმებული მასალა, დიალექტური ტექსტი, სხვადასხვა სამეცნიერო წყარო... ზოგიერთი სალექსიკონო სტატია მხოლოდ ერთი წყაროს მონაცემებზე დაყრდნობით არის შექმნილი, ზოგი კი რამდენიმე წყაროდან გამოკრებილი ინფორმაციითაა გამდიდრებული.

მას შემდეგ, რაც შეიქმნა ჩვენთვის საინტერესო ლექსიკონების ონლაინვერსია, სპეციალურად ამ მიზნით დაწერილი პროგრამული ინსტრუმენტის საშუალებით შეიქმნა სალექსიკონო სტატიის მაკეტი – ლემის, ლემის ლათინური ტრანსკრიპციის, ლემის გრამატიკული მარკერის, ვარიაციებისა და მათი გრამატიკული მარკერების, სიტყვის წარმომავლობისა და უცხო ენაზე დაწერილობის, ილუსტრაციის, ენციკლოპედიური ინფორმაციის, აგრეთვე ლექსიკოგრაფიული ბმულების (იგივეა რაც, იხილე) გადმოცემის სასურველი გრაფიკული „ესკიზი“. ამ პროგრამული ხელსაწყოს საშუალებით შეიქმნა ლექსიკონების ბეჭდური პირველსახე, რომელსაც საბოლოო სასტამბო ვერსიის მისაღებად დასჭირდა მხოლოდ შრიფტის შეცვლა, სპეციფიკური ასონისმების გადამოწმება და ზედაპირული კორექტურა.

ფაქტობრივად, კორპუსული ლექსიკონების მომზადების მთლიანი ციკლის „გამოცდამ“ წარმატებით ჩაიარა. როგორ გამოიყურება ბაზიდან გენერირებული ფერეიდნული ლექსიკონის ერთი გვერდი, იხილეთ სურათზე (სურ.1).

ამ ტექსტში მხოლოდ შრიფტია შეცვლილი. ყველაფერი დანარჩენი (ლემისა და ვარიაციების გამუქება, სტილი, ინტერვალი, მარკერის აღნიშვნა, ვარიაციები და მათი მარკერები, ლათინური ტრანსკრიპცია, ილუსტრაციის სტილი (დახრილი), ილუსტრაციებს შორის სასვენი ნიშნები... ავტომატურად არის გენერირებული.

ამ ექსპერიმენტმა შექმნა ლექსიკოგრაფიული სამუშაოს დისტანციურად ორგანიზების პერსპექტივა.

სამივე ლექსიკონზე დაახლოებით 10 ადამიანი მუშაობდა სხვადასხვა სტატუსით (შემდგენლები, ოპერატორები, კორექტორები, რედაქტორები) და ისინი ხელს არ უშლიდნენ ერთმანეთს. საგამომცემლო ვერსიის მომზადება დაიწყო 2018 წლის აპრილში და დამთავრდა სარეკორდო დროში – 2018 წლის აგვისტოს ბოლოს.

*ქედი შაბის, ერთმა აბრეს კისა შაბის, ერთმა კალაოტ საოზნი შაბის, შითამ შომოლოართყე ემ ახანსლუსა; ეხლა პარ სახლნი თითან აბრეს იყო. სუ სახლსლა აბრესლა აქლა; ეხლა ადარ არილა ის აბრეს-ბილა.*

**აბოხავა** (N) abohava სპარსული **ابو و ابه** āb-o-havā. კლიმატი. აღსამდგომელი ნამა აბოხავა ვოქ. ბუჰარანი აბოხავა ჩონ ვოქ.

**აბრეშამი** (N) abrešami სპარსული **ابرهشام** abrišām. აბრეშუმი.

**აბრი** (N) abri ზღაპრული ცხენის სახელი. პარსული ზღაპარი ყოფილა, ცხენის სახელი აბრი იყო. ემსწილებს ქართულად უამბიანეც ემ სახელით.

**აბრი** (N) abri სპარსული **ابره** abr. ღრუბელი.

*ცილი სათი იყო /თორღელი/.*

**აბუეზბა** (Msd) abuzeba იბუეზბის ibuzebis (V:Sg.3.Prs), **აბუეილა** abuzila (V:Sg.3.Result I). სიცივისგან კარკალი. ეს ემსწილი, მიხე, როგო იბუეზბის; კოჭლი შამად ალიას არცრა არ აცია, სიცივით აბუეილა და ჩემ ამხანავიც ძველსავით აკეულებია.

**აბქეში** (N) ) abkeši სპარსული **ابکشی** ābkeš. საწური. ბრინჯი აბქეში ვეცნ.

**აბქეში** (N) abkeši სპარსული **ابکشی** ābkeš. ის, **დამფოსტი**, წყალში მოშალეზული ბრინჯი, მაკარონი. ვაფარჩევეთ და ვაგრეცხეთ და წყალს დავასხამთ, მარილს ავადლეუთ და აბქეშს ვიქო და დამზე დავადკამთლა და ავე ზად ბრინჯი მოვლა რა ბრინჯი აბქეში, მაკარონი აბქეში; ავადლეუთ

**სურ. 1.**

ამ ლექსიკონების გამოცემა იყო ქდკ-ის ლექსიკოგრაფიული კონცეფციის ბოლო ტექნოლოგიური ეტაპი, რომელიც წარმატებით განხორციელდა.

ქდკ-ის ლექსიკოგრაფიული კონცეფცია არაერთხელ წარმოვადგინეთ როგორც პროექტის დასაწყისში, ისე მისი განხორციელების სხვადასხვა ეტაპზე (ბერიძე და ნადარაია 2018; Beridze and Nadaraia 2009, 2016; Beridze, Nadaraia and Lordkipanidze 2015a, 2015b; Beridze, Bakuradze and Nadaraia 2017). ეს კონცეფცია გზადაგზა იხვეწებოდა და იცვლებოდა. თუმცა მთავარი იდეა უცვლელი დარჩა: კორპუსის ლექსიკოგრაფიული ბაზა მის ტექსტურ, საბიბლიოთეკო და კარტოგრაფიულ კომპონენტთან ერთად უნდა იქცეს ენობრივი მონაცემების დამუშავების ერთიან ტექნოლოგიურ ინსტრუმენტად – **ენის ციფრული კვლევის პროცესორად**, რომელიც გააერთიანებს კორპუსის, ლექსიკონებისა და მონაცემთა კარტოგრაფიული ვიზუალიზაციის სამომხმარებლო პლატფორმებს.

ქდკ-ის არსებული ბაზა ამგვარი პროცესორის პროტოტიპულ მოდელად შეიძლება ჩაითვალოს.

ლექსიკოგრაფიული რედაქტორი ერთ-ერთი ყველაზე რთული და ამასთან, კარგად დამუშავებული სისტემაა ამ მოდელში. ის ერთდროულად არის ერთიანი მრავალდარგოვანი ლექსიკოგრაფიული ბაზის შექმნის ხელსაწყო და ამასთანავე, ამ ბაზიდან სხვადასხვა ტიპის ლექსიკონების გენერირების საშუალება. დღეს ამ მრავალფუნქციური ლექსიკოგრაფიული სისტემის გამოყენება მხოლოდ პროექტის სამუშაო ფაზისთვისაა შესაძლებელი, თუმცა მისი პროცესორად გარდაქმნის იდეა გულისხმობს სამომხმარებლო ფუნქციების განვითარებასა და მისი ფართო ხელმისაწვდომობის უზრუნველყოფას.

## 2. ქდკ-ის ლექსიკოგრაფიული ბაზის სტრუქტურა

### 2.1. ქდკ-ს ლექსიკონის საინფორმაციო ველების აღწერა

ლექსიკოგრაფიული ბაზის ტექნოლოგიური ჩარჩო – ქდკ-ის ლექსიკოგრაფიული რედაქტორი არის ხელსაწყო, რომელიც ლექსიკოგრაფიული ბაზის ინფორმაციის „შენახვისა“ და მისი სტრუქტურირების საშუალებას იძლევა. ამასთან ის არის ამ ინფორმაციის ძიების ეფექტური ხელსაწყო.

რედაქტორი წარმოდგენილია **სამომხმარებლო** და **სამუშაო** ნიღბების (ე.წ. ინტერფეისების) სახით.

**ლექსიკოგრაფიული რედაქტორის სამუშაო ნიღაბი** მხოლოდ ავტორიზებული მომხმარებლისთვისაა ხელმისაწვდომი ბაზასთან დაშვების სხვადასხვა ხარისხით (ადმინისტრატორები, ოპერატორები, რედაქტორები – მათ სხვადასხვა მასალაზე და სხვადასხვა ხარისხის წვდომა აქვთ უსაფრთხოების დაცვის მიზნით).

ლექსიკოგრაფიული ბაზის ფორმირება სწორედ სამუშაო ნიღბის საშუალებით არის შესაძლებელი.

**ლექსიკოგრაფიული ინფორმაციის დამატება** რედაქტორში შესაძლებელია როგორც პირდაპირ სისტემის შევსების გზით, ისე წინასწარ მომზადებული სტრუქტურირებული ცხრილების ატვირთვით. ხელსაწყო არ ითხოვს მომზადებული ფაილების რაიმე სხვა ფორმატს და, შესაბამისად, მასზე მომუშავე ლექსიკოგრაფების ან ოპერატორების განსაკუთრებულ სპეციალურ ტექნიკურ კვალიფიკაციას.

ლექსიკოგრაფიული რედაქტორის სამუშაო „სივრცე“ მოიცავს სხვადასხვა სარედაქციო ველებს და ფუნქციონალებს:

**სალექსიკონო სტატია** რედაქტორში სტრუქტურირდება შემდეგი ველების სახით:

**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.

**შენიშვნის ველი** – ეს ველი სამუშაო ჯგუფის წევრებს შორის ინფორმაციის გაცვლისთვისაა: აქ იწერება რედაქტირების ხარვეზის ან რომელიმე დეფექტური ველის შესახებ ინფორმაცია, რომელიც გათვალისწინებულ უნდა იქნეს სტატიის გამოქვეყნებამდე. რადგან სალექსიკონო სტატიას ხშირად რამდენიმე ადამიანი ამუშავებს ერთდროულად, ეს ველი ამ სამუშაოს ერთგვარ ისტორიას ქმნის.

**შენიშვნა:** შენახვისა და შენიშვნის ველები ლექსიკოგრაფიული ბაზის „კულისებია“ და საძიებო სისტემაში ისინი ჩართული არ არიან.

**ლემის დამატების ველი** – ლემა ამ ველში აიტვირთება, როგორც ვთქვით, წინასწარ მომზადებული სიის სახით ან ხელით ჩაიწერება. ლემათა სია ან ბეჭდური ლექსიკონების ციფრული ვერსიიდან არის ფორმირებული, ან ტექსტების კორპუსიდან გენერირებული. ტექსტების კორპუსიდან სიტყვანის გენერირება სამუშაო ნილაბში ჩაშენებული სპეციალური ფუნქციონალითაა შესაძლებელი.

**შესატყვისის დამატების ველი** – ამ ველში მხოლოდ მარტივი შესატყვისი იწერება, თუკი დიალექტურ ლემას ასეთი შესატყვისი გააჩნია სალიტერატურო ენაში. ორზე მეტი შესატყვისი იწერება ერთ უჯრაში და ერთმანეთისგან გამოიყოფა მძიმით.

**განმარტების დამატების ველი** – ამ ველში იმ დიალექტურ ფორმათა განმარტება იწერება, რომელთაც არ გააჩნიათ შესატყვისი სალიტერატურო ენაში; ზოგჯერ შეიძლება შესატყვისის ველიც შეივსოს და განმარტებისაც.

**ენციკლოპედიური ინფორმაციის დამატების ველი** – ამ ველში შესაძლებელია ყველანაირი ინფორმაციის დამატება: ეთნოგრაფიული, ეტიმოლოგიური, ისტორიული და ა. შ. ამ ველის შევსება აუცილებელი არ არის, მაგრამ მისი საშუალებით მომავალში შესაძლებელია ერთიანი ლექსიკოგრაფიული ბაზიდან ენციკლოპედიური ლექსიკონების მობილიზება.

**სიტყვის უცხოური წარმომავლობის შესახებ ინფორმაციის დამატების ფუნქციონალი** (რომელი ენიდანაა ნასესხები) – არჩევითი სიიდან საჭირო ენის არჩევით დამატება.

**სიტყვის უცხოური დაწერილობის დამატების ველი** (როგორ იწერება სიტყვა იმ ენაზე, საიდანაც არის ნასესხები).

**ინფორმაცია სიტყვის სტრუქტურის შესახებ** (ენკლიტიკა, კომპოზიცია, მრავალკომპონენტური სიტყვა, რთული გრამატიკული ფორმა, მარტივი შესიტყვება).

**ინფორმაცია ლექსიკური ტიპის შესახებ** (ნეიტრალური, ტროპული, დარგობრივი).

**ინფორმაცია სემანტიკური კლასის შესახებ** – არჩევითი სიიდან სემანტიკური ტიპის არჩევა.

**ინფორმაცია ლექსიკოგრაფიული წყაროს / წყაროების შესახებ** – არჩევითი სიიდან საჭირო წყაროს დამატება.

**ლემის გრამატიკული მარკერი** – არჩევითი სიიდან საჭირო მარკერის დამატება.

**ლემის გრამატიკული (ფორმაწარმოებითი) ვარიაცია** – სპეციალური ფუნქციონალი, რომელიც საშუალებას იძლევა, დაემატოს მოცემული ლემის ნებისმიერი რაოდენობის პარადიგმული რეალიზაცია, აგრეთვე, გრამატიკულ მარკერთა ორდონიანი იერარქიული სიიდან დაემატოს ნებისმიერი რაოდენობის მახასიათებელი.

**ლემის სიტყვაწარმოებითი ვარიაცია** – სპეციალური ფუნქციონალი, რომელიც საშუალებას იძლევა, დაემატოს მოცემული ლემის ნებისმიერი რაოდენობის სიტყვაწარმოებითი ვარიაცია (მიმღეობა, მასდარი და სხვ.), აგრეთვე, საჭიროების შემთხვევაში გრამატიკულ მარკერთა ორდონიანი იერარქიული სიიდან დაემატოს ნებისმიერი რაოდენობის მახასიათებელი.

**ლემის ფონეტიკური ვარიაცია** – სპეციალური ფუნქციონალი, რომელიც საშუალებას იძლევა, დაემატოს ლემის ნებისმიერი რაოდენობის ფონეტიკური ვარიანტი.

**მითითება სინონიმზე (იგივეა რაც)** – მითითების ეს ფუნქციონალი საშუალებას იძლევა, მიეთითოს ყველა ნაცნობ სინონიმზე როგორც მოცემული დიალექტის ფარგლებში, ისე მთელი ლექსიკოგრაფიული მასივის მასშტაბით.

**მითითება სხვა ლექსიკურ ერთეულზე (იხილეთ)** – მითითების ეს ფუნქციონალი საშუალებას იძლევა, მიეთითოს მოცემულ ლემასთან მნიშვნელობით, ძირით, ფორმით ან სხვა რამ საინტერესო ნიშნით დაკავშირებულ ფორმაზე; მითითება შეიძლება როგორც ერთი დიალექტის ლექსიკონის შიგნით, ისე მთელი ბაზის მასშტაბით. ფუნქციონალი საშუალებას იძლევა, ამოვკრიბოთ ბაზიდან ყველა მსგავსი ძირის სიტყვა, გადავამოწმოთ კავშირი და ყველა ერთი ძირის სიტყვა დავაკავშიროთ ერთმანეთთან ბმულით: **იხილეთ**.



**ილუსტრაცია** – ილუსტრაციის დამატების ველში ემატება სასურველი რაოდენობის ილუსტრაციები.

**ილუსტრაციის თარგმანი ან კომენტარი** – ველი ივსება საჭიროების შემთხვევაში.

**ილუსტრაციის წყარო** – ველი ივსება საჭიროების შემთხვევაში.

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

რედაქტორში შესაძლებელია რამდენიმე (შეუზღუდავი რაოდენობის) პირის ერთდროულად მუშაობა. მეტიც, შესაძლებელია ერთი და იმავე ლექსიკონის სხვადასხვა სტატუსის მქონე სტატიაზე ან ერთი და იმავე სალექსიკონო სტატიის სხვადასხვა ველის შევსებაზე / რედაქტირებაზე სხვადასხვა ადამიანის მუშაობა, ისე, რომ ისინი ხელს არ უშლიდნენ ერთმანეთს. მაგალითად, შესაძლებელია, რედაქტორმა იმუშაოს მხოლოდ იმ სტატიებზე, რომელსაც აქვს სტატუსი „რედაქტირებადი“ – საბოლოოდ შეამოწმოს და მიანიჭოს ან არ მიანიჭოს სტატუსი „გამოქვეყნებული“. რაც შეეხება, ლექსიკოგრაფთა და ოპერატორთა ჯგუფს, ისინი მუშაობენ იმ სტატიებზე, რომელთა სტატუსია – „ახლად შექმნილი“. შესაძლოა, სხვადასხვა პირი მუშაობდეს ერთი სტატიის სხვადასხვა ველის შევსებაზე და ა. შ.

„ახლად შექმნილი“ სტატიები წარმოადგენს ხან მხოლოდ ლემას, ხან ლემასა და მის თანმხლებ არასტრუქტურირებულ ინფორმაციას (განმარტება, ილუსტრაცია, კონტექსტები...), რომელიც ან ხელით შეჰყავს ოპერატორს, ან ნახევრადავტომატურად არის ატვირთული რედაქტორში, ან სპეციალური ხელსაწყოთი არის „გადმოზიდული“ ტექსტური კორპუსიდან.

ეს ინსტრუმენტი ქდკ-ის ლექსიკონების ადრეულ ეტაპზე გამოქვეყნებისა და მათი მუდმივი სრულყოფის საშუალებას იძლევა. მაგალითად, თავდაპირველად გამოვაქვეყნეთ ფერეიდნული ლექსიკონი ასეთი სახით: სალექსიკონო სტატია მოიცავდა მხოლოდ ლემას, მის განმარტებას და ილუსტრაციას. საწყის სტატიებს კი თანდათან დავემატა: უცხოური წარმომავლობის, გრამატიკული და ფონეტიკური ვარიაციების, სიტყვის სტრუქტურის, სემანტიკური ჯგუფის, სიტყვის სოციალური სტატუსისა და სხვ. შესახებ ინფორმაცია, ენციკლოპედიური ცნობები და სხვ.



სურ.2 გამოქვეყნებული სტატია

**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 სალექსიკონო სტატია ხელმისაწვდომი ხდება მომხმარებლისთვის (სურ.2).

## 2.2. ლემის წარმოდგენის სახე ქდკ-ის ლექსიკონებში

**ლემად** შეიძლება წარმოდგენილი იყოს როგორც ცალკე სიტყვა, ასევე მყარი შესიტყვება – მრვალკომპონენტიანი ერთეული, რომელიც ფაქტობრივად, განიხილება, როგორც ერთი სიტყვა. მყარი შესიტყვება შეიძლება შეიცავდეს სიტყვათა შეუზღუდავ რაოდენობას (მთელ წინადადებასაც კი – ფრაზეოლოგიზმის შემთხვევაში).

**ლემის ველში** შეიძლება ჩაიწეროს მხოლოდ ერთი ერთეული – სიტყვა ან შესიტყვება, ყოველგვარი ვარიაციულობის აღნიშვნის გარეშე.

**სახელები** – წარმოდგენილია მხოლოდითი რიცხვის სახელობითი ბრუნვის ფორმით.

**ზმნა** – ძირითადად წარმოდგენილია მხოლოდითი რიცხვის აწმყოს მესამე პირის ფორმით, თუმცა გარკვეულ შემთხვევებში (ლექსიკოგრაფიული წყაროს გათვალისწინებით) შესაძლოა წარმოდგენილი



- დაემატება ფონეტიკური ვარიანტები (ფონეტიკური ვარიანტების რაოდენობა არ არის შეზღუდული)

- დაემატება გრამატიკული / სიტყვაწარმოებითი ვარიანტები (ვარიანტების რაოდენობა არ არის შეზღუდული).

ლემასაც და გრამატიკულ / სიტყვაწარმოებით ვარიაციებსაც მიეწერება გრამატიკული ინფორმაცია. გრამატიკული მარკერები დაემატება გრამატიკული მარკერების იერარქიული სიებიდან. პირველი იერარქიის მარკერთა სიაში მოცემულია მეტყველების ნაწილების ჩამონათვალი, აგრეთვე მასდარის, მიმღეობის და ზოგიერთი ფრაგმენტული ელემენტის (სახელური ფუძე, ზმნური ფუძე, ზმნისწინი...) მარკერი, რომლებითაც მოინიშნება ლემა.

მეორე იერარქიის მახასიათებლებში მოცემულია გრამატიკული კატეგორიების ძირითადი მახასიათებლები, აგრეთვე ზოგი სიტყვაწარმოებითი მახასიათებელი. მახასიათებლები შერჩეულია საერთაშორისო სტანდარტების გათვალისწინებით

**ლემა** მხოლოდ ერთი – ზედა იერარქიის მარკერით მოინიშნება.

**ფონეტიკური ვარიანტი** – არ მოინიშნება გრამატიკული მარკერებით. მისი გრამატიკული სტატუსი ლემის სტატუსის იდენტურია.

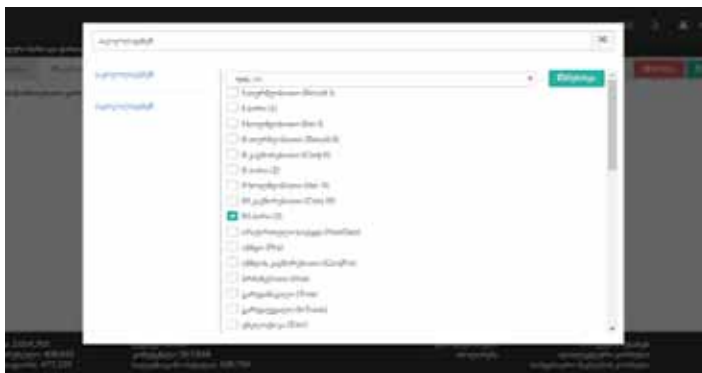
**გრამატიკული / სიტყვაწარმოებითი ვარიანტი** – მოინიშნება როგორც გრამატიკული კლასის, ისე გრამატიკული კატეგორიის მახასიათებლებით (სურ. 4).

ყოველივე ამის შედეგად, როგორც აღვნიშნეთ, ონლაინლექსიკონის სიტყვანში ერთი ლემა რამდენიმე სიტყვაფორმითაა წარმოდგენილი. თითოეულ მათგანს ერთი სალექსიკონო სტატია შეესაბამება.

ონლაინლექსიკონების სამომხმარებლო ბლოკში შესაძლებელია ძიება ფილტრებით: მხოლოდ ლემის ძიება, ლემისა და სიტყვაფორმის ძიება; სრული სიტყვის ან სიტყვის ფრაგმენტით (იწყება, მთავრდება, შეიცავს); გრამატიკული კლასის მარკერით ძიება; გრამატიკული კატეგორიის მარკერით ძიება.

#### 2.4. სალექსიკონო სტატიის ინფორმაციული ბლოკი (ლექსიკონის „მარჯვენა“ მხარე)

როგორც ზემოთ აღვნიშნეთ, განმარტებითი ინფორმაცია სიტყვის შესახებ ქდკ-ის ლექსიკონში სამ სხვადასხვა ველში ნაწილდება. ესენია: მარტივი შესატყვისი, სიტყვის არსის ახსნა, განმარტება, ენციკლოპედიური ინფორმაცია.



სურ.4 გრამატიკული ინფორმაციის დამატება მახასიათებელთა იერარქიული სიიდან

რატომ გავყავით მარტივი შესატყვისისა და განმარტებითი ინფორმაციის ველები? – ამით შევქმენით შესაძლებლობა, რომ მომავალში, თუკი პროექტი ამ მიმართულებით განვითარდება, მარტივი შესატყვისის ველში ჩაწერილი სიტყვები ავტომატურად დავაკავშიროთ სალიტერატურო ენის განმარტებით ან ქართული ენის ნებისმიერ ორენოვან ლექსიკონთან – სალიტერატურო შესატყვისის ველის ინფორმაციას ამ შემთხვევაში ავტომატურად მივებმება უცხოენოვანი შესატყვისი და სათარგმნი დარჩება მხოლოდ მეორე და მესამე განმარტებითი იერარქიის მასალა.

სამომხმარებლო ვერსიაში ეს ინფორმაცია მარკირებული არ არის. განმარტების ამ სამ ველს მხოლოდ სამუშაო ნილაბი განარჩევს. ის საშუალებას იძლევა, ველები სხვადასხვა დროს / ან პარალელურად სხვადასხვა ადამიანის მიერ შეივსოს, უბრალო განმარტებითი დიალექტური ლექსიკონი ენციკლოპედიურ ლექსიკონად იქცეს. უცხო ენაზე ინფორმაციის დამატების შემთხვევაში – ორენოვან ლექსიკონად და ა. შ.

**სხვა ლექსიკოგრაფიული ინფორმაცია** – ილუსტრაციებისა და ლექსიკოგრაფიული წყაროების მითითების ფუნქციონალი საშუალებას იძლევა, დავამატოთ ნებისმიერი რაოდენობის ილუსტრაცია და საჭიროების შემთხვევაში თითოეულ ილუსტრაციას მივუწეროთ კომენტარი ან თარგმანი, აგრეთვე ილუსტრაციის წყარო.

ლექსიკოგრაფიული წყაროს „არჩევა“ ხდება ბიბლიოგრაფიის ბაზიდან, რომელიც წინასწარ არის შემუშავებული და აღჭურვილია სტანდარტული მეტამახასიათებლებით.

ლექსიკონის რედაქტორის ეს ფუნქცია საშუალებას გვაძლევს, პირველწყარო ლექსიკონი გარდავექმნათ სრულიად ახალი სახის ლექ-

სიკონად, შევავსოთ მისი „ნაკლული“ ველები ან გავამდიდროთ საილუსტრაციო მასალით და ამასთან მივუთითოთ ყველა ლექსიკოგრაფიული წყარო, რომელიც იქნა გამოყენებული ამ პროცესში. ამ გზით ჩვენ ვაერთიანებთ ხან სამ-ოთხ სხვადასხვა ლექსიკოგრაფიულ წყაროს, მათ შორის საკუთარი ექსპედიციების დროს მოპოვებულ მასალებსაც. ყველა ლექსიკოგრაფიული წყარო ებმის თითოეულ სალექსიკონო სტატიას და არა მთელ ლექსიკონს. საჭიროების შემთხვევაში შესაძლებელია მხოლოდ სასურველი წყაროს (ავტორის) სალექსიკონო სტატიების ძიება.

ილუსტრაციის წყაროს ველში შეიძლება წარმოდგენილი იყოს როგორც ბიბლიოგრაფიული, ისე პიროვნული ან გეოგრაფიული მახასიათებლები.

**დამატებითი ლინგვისტიკური ინფორმაცია სიტყვის შესახებ** – გარდა ლემის სხვადასხვა ტიპის ვარიაციებისა, რაზეც უკვე გვქონდა საუბარი, ლექსიკონში წარმოდგენილია სხვა ტიპის მდიდარი ლინგვისტიკური ინფორმაციაც, როგორცაა:

**უცხო სიტყვა და უცხო სიტყვის დაწერილობა** – ეს ინფორმაცია ორი სხვადასხვა ოპერაციით დაემატება: ენების ჩამონათვალიდან აირჩევა სესხების ენა და სპეციალურ ველში მიეწერება შესაბამისი სიტყვა სესხების ენაზე (სესხების ენა ყოველთვის არ ემთხვევა სიტყვის ისტორიულ წარმომავლობას).

უცხოენოვანი ლექსიკის მარკირება ამ დიალექტებში, მათი ქართული (დიალექტური) სინონიმების მითითება ან ზუსტი ქართული შესატყვისების მიწერა, ჩვენს ონლაინლექსიკონებს სძენს დამატებით ფუნქციას – ეს არის დიალექტურ დიასპორებში ქართული ენის სწავლების ფუნქცია. სამომხმარებლო ძიება ზუსტად აჩვენებს ყველა ნასესხებ სიტყვას, მათ ყველა დიალექტურ პარალელურ შესატყვისს (ბმა: „იგივეა რაც“) და ქართულ შესატყვისს ან განმარტებას.

**სალექსიკონო რედაქტორში აგრეთვე შესაძლებელია:** ლექსიკურ/სემანტიკური ჯგუფების მიხედვით მარკირება; სიტყვის სტრუქტურის მიხედვით მარკირება; რთული ლემის ფუნქციურ-სემანტიკური მარკირება.

**სიტყვის (ლემის) სტრუქტურის საკლასიფიკაციო სიის დაქვემდებარებული იერარქიული სიაა ლექსიკური ტიპის მარკერების ჩამონათვალი, რომელიც საშუალებას მოგვცემს როგორც მარტივი, ისე რთული სტრუქტურის მქონე ენობრივი ერთეულების კლასიფიკაცია მოვახდინოთ ლექსიკური ტიპის მიხედვით.**

**ლექსიკური ტიპის მახასიათებლებში** გამოვყავით: ზოგადი ლექსიკა (ლექსიკური ფონდი ტროპული ლექსიკისა და სპეციფიკური

ლექსიკის გამოკლებით), ტროპული ლექსიკა და სპეციფიკური / დარგობლივი ლექსიკა;

სამივე ეს ჯგუფი, თავის მხრივ, ექვემდებარება იერარქიულ განდასებას:

**ტროპულ ლექსიკაში** გამოიყოფა: ანდაზები, ფრთიანი გამოთქმები, სხვადასხვა ფორმულები (წყევლის, ლოცვის, გინების...)

**ტერმინებში** – მიეთითება გამოყენების დარგები.

**ნეიტრალური ლექსიკიდან** გამოიყოფა საკუთარი სახელები (ანთროპონიმები, ჰიდრონიმები, გვარები, ტოპონიმები, ზოონიმები) და სხვა სემანტიკური ჯგუფები.

**შიდა მიმართებები** ქდკ-ის ლექსიკონებში აღინიშნება ბმულებით: „იგივეა რაც“ და „იხილე“. პირველი ბმული აკავშირებს სინონიმურ სიტყვებს როგორც ერთი დიალექტის ლექსიკონში, ისე მის „გარეთ“ – სხვა დიალექტების მასალაში. მეორე კი აგზავნის სხვა ნიშნით (ფორმობრივი, სტრუქტურული, ფრაზეოლოგიურ კომპონენტზე ორიენტირებული და სხვ.) დაკავშირებულ ფორმებთან.

შიდა ბმულები ორმხრივი ხასიათისაა და ერთხელ მონიშნული მიმართება ავტომატურად აღინიშნება საპირისპირო მიმართულებით.

### ერთი ცნება – ერთი მნიშვნელობა

იმის გამო, რომ ქდკ-ის ლექსიკონები იქმნება კორპუსის პარალელურად და გამოიყენება კორპუსული მასალის ნაწილობრივი ავტომატური ანალიზის პროცესშიც, თავიდანვე მივიღეთ გადაწყვეტილება, რომ სიტყვის ყველა მნიშვნელობა (ომონიმური, გადატანითი (ტროპული), ტერმინოლოგიური, ოკაზიური...) გაგვეტანა ლექსიკონში ცალკე სალექსიკონო ერთეულად. თითოეულ ასეთ მნიშვნელობაზე დგება ცალკე სალექსიკონო სტატია და მყარდება ძირითად სტატიაზე მიმართება ბმულით „იხილე“.

### მორფოსინტაქსური და სემანტიკური მარკერები

ქდკ-ისთვის მარკერები შემუშავდა არსებული სტანდარტებისა (TEI, EAGLES, The Leipzig Glossing Rules) და ქართული ეროვნული კორპუსის (GNC: <http://gnc.gov.ge>) მონაცემების გათვალისწინებით. ქართული დიალექტების სპეციფიკისა და ჩვენი ზოგიერთი კონკრეტული ამოცანის შესაბამისად მოხდა მარკერთა ნაწილობრივი ჰარმონიზაცია. დავამატეთ რამდენიმე აუცილებელი მარკერი.

როცა განსაკუთრებულ ამოცანებზე ვსაუბრობთ, მაგალითად ისევ საზღვარგარეთ გავრცელებულ დიალექტებს, კერძოდ, ფერეიდნულ დიალექტს მოვიყვანთ. ეს ენობრივი კოლექტივი ოთხი საუკუნის განმავლობაში იმყოფებოდა ენობრივი კუნძულისათვის დამახასიათებელ იზოლაციაში. ამის წყალობით შეინარჩუნეს იქაურმა ქართველებმა მშობლიური ენა. გასული საუკუნის მეორე ნახევრიდან კუნძულის იზოლაცია დაირღვა და ის სპარსული ენის ძლიერ ზეგავლენაში მოექცა. ამავ პერიოდიდან ინტენსიური გახდა ფერეიდანის მოსახლეობის კავშირები ისტორიულ სამშობლოსთან. გასული საუკუნის ბოლოდან შეინიშნება კულტურული და ენობრივი რეინტეგრაციის ინტენსიური პროცესი. ამ პროცესების აღსაწერად და დასაკვირვებლად როგორც ფერეიდანის, ისე სხვა მსგავსი ერთეულების მეტყველებაში, შემოვიღეთ ახალი მარკერები:

Pseudo – სალიტერატურო ქართულის მცდარი აღქმის შედეგად შექმნილი ფორმები, რომლებიც დასტურდება ქდკ-ის ტექსტებში.

New – სალიტერატურო ქართულის გავლენით დიალექტური ფორმის გვერდით გაჩენილი ლექსიკური პარალელიზმები.

მორფოსინტაქსური და სემანტიკური მარკირება ლექსიკონებში ორსაფეხურიანი იერარქიული სიებით ხდება: პირველი იერარქიის სიაში ლემის მარკერებია მოცემული (გრამატიკული კლასის ნიშნები და ზოგიერთი სხვა სალექსიკონო ერთეულის აღსანიშნი მარკერები: თანდებული, ზმნისწინი და ა. შ.). მეორე იერარქიის სიებში კი გვაქვს მორფოლოგიური კატეგორიების აღმნიშვნელი მარკერები, ზოგი სიტყვაწარმოებითი მარკერი, სემანტიკური მარკერები). ყველა ეს მარკერი სიიდან არჩევისა და დამატების წესით მიეწერება სათანადო სალექსიკონო ერთეულს ან აისახება სალექსიკონო სტატიის ცალკე ველში.

### **ძიება ლექსიკონებში**

ქდკ-ის სტრუქტურირებული ლექსიკონები უკვე იძლევა საშუალებას, მის სამუშაო და სამომხმარებლო ინტერფეისებში განვახორციელოთ საკმაოდ რთული დიფერენცირებული ძიება: ძიება სიტყვის ფრაგმენტის მიხედვით, ლემის მიხედვით ძიება / ლემისა და ვარიაციების მიხედვით ძიება, უცხო სიტყვის ძიება / კონკრეტული უცხო ენიდან შემოსული სიტყვის ძიება, დიალექტის მიხედვით ძიება – ყველა დიალექტში ერთდროულად / რომელიმე ერთ დიალექტში ძიება, მეტყველების ნაწილის მიხედვით ძიება, გრამატიკული მახასიათებლის მიხედვით ძიება, სალექსიკონო სტატიის სტატუსის მიხედვით ძიება



(საწყისი სტატია, რედაქტირების სტადიაში მყოფი სტატია, დასრულებული სტატია), ლექსიკოგრაფიული წყაროს მიხედვით ძიება...

### 3. განვითარების პერსპექტივა

პროექტის განვითარების პერსპექტივად გვესახება ქართული დიალექტური კორპუსის ბაზაზე შექმნილი მრავალფუნქციური კვლევითი და ინფორმაციული სისტემის, როგორც პროტოტიპული მოდელის, მიხედვით ისეთი უნივერსალური პროცესორის შექმნა, რომელიც იქნება ენობრივი მონაცემების კორპუსული, ლექსიკოგრაფიული და კარტოგრაფიული დამუშავების ინსტრუმენტი და მისაწვდომი იქნება თითოეული დაინტერესებული პირისთვის, უკვე არა როგორც მხოლოდ სამომხმარებლო პროდუქტი, არამედ, როგორც პროგრამული საშუალება საკუთარი პროექტების განსახორციელებლად. ეს მიზანი დაედო საფუძვლად ჩვენს ახალ სამომავლო პროექტს: **ენის ციფრული კვლევის პროცესორი: კორპუსები, ლექსიკონები, რუკები.**

### დამოწმებანი

- ბერიძე, მარინა, ლია ბაკურაძე და ნიკოლოზ ნახუცრიშვილი. 2018. *ფერეიდნული დიალექტის კორპუსული ლექსიკონი*. Wojnowice: KEW – Kolegium Europy Wschodniej.
- ბერიძე, მარინა, ნინო შარაშენიძე და ელენე კიკნაძე. 2018. *იმერული დიალექტის კორპუსული ლექსიკონი*. Wojnowice: KEW – Kolegium Europy Wschodniej.
- ბერიძე, მარინა და დავით ნადარაია. 2018. დიდი ქართული დიალექტური ბაზა – პრობლემები და პერსპექტივები. საერთაშორისო სამეცნიერო კონფერენციის „ჰუმანიტარული მეცნიერებები ინფორმაციულ საზოგადოებაში-III“ შრომების კრებული, ბათუმი: შოთა რუსთაველის სახელმწიფო უნივერსიტეტის გამომცემლობა.
- ცოცანიძე, გიორგი და მარინა ბერიძე. 2018. *თუშური დიალექტის კორპუსული ლექსიკონი*. Wojnowice: KEW – Kolegium Europy Wschodniej.
- Beridze, Marina, Lia Bakuradze and David Nadaraia. 2017. Georgian Dialect Corpus: Linguistic and Encyclopedic Information in Online Dictionaries. *Journal of Linguistics*, vol. 68, no. 2, 109-121.
- Beridze, Marina and David Nadaraia. 2016. Lexicographic Potential of the Georgian Dialect Corpus. In *Proceedings of the XVII EURALEX Interna-*

*tional Congress, Lexicography and Linguistic Diversity*, September 6-10, 2016, eds. Tinatin Margalitadze and George Meladze, 300-309. Tbilisi: Ivane Javakhishvili Tbilisi State University Press.

Beridze, Marina, Liana Lordkipanidze and David Nadaraia. 2015a. The Georgian Dialect Corpus: Problems and prospects. In *Historical Corpora Challenges and Perspectives*, eds. Jost Gippert and Ralf Gehrke, 323-333. (The volume contains 23 papers read at the international conference “Historical Corpora 2012”, hosted by the LOEWE Research Cluster “Digital Humanities” of the State of Hesse at the University of Frankfurt on December 6-8, 2012.) Tübingen: Narr.

Beridze, Marina, Liana Lordkipanidze and David Nadaraia. 2015b. Dialect Dictionaries in the Georgian Dialect Corpus. In *Proceedings of the 10th International Symposium on Logic, and Computation*, TbiLLC 2013, held in Gudauri, Georgia, September 23-27, 2013. Revised Selected Papers, XIV, eds. Martin Aher, Daniel Hole, Emil Jeřábek, Clemens Kupke, 82-96. Berlin: Springer.

Beridze, Marina and David Nadaraia. 2009. The Corpus of Georgian Dialects. In *Proceedings of the NLP, Corpus Linguistics, Corpus Based Grammar Research*. The Fifth International Conference held in Smolenice, Slovakia, 25-27 November 2009, Eds. Jana Levická and Radovan Garabík, 25-34. Bratislava: Tribun.