

ივანე ჯავახიშვილის სახელობის თბილისის სახელმწიფო უნივერსიტეტი  
ჰუმანიტარულ მეცნიერებათა ფაკულტეტი  
ინგლისური ფილოლოგიის დეპარტამენტი  
ლექსიკოგრაფია

სოფიკო დარასელია

ქართულ-ევროპული ტიპის ახალი სასწავლო ლექსიკონების შედგენის  
საკითხები კორპუსული მეთოდოლოგიის გამოყენებით

ფილოლოგიის დოქტორის (Ph.D.) აკადემიური ხარისხის მოსაპოვებლად  
წარმოდგენილი

დ ი ს ე რ ტ ა ც ი ა

სამეცნიერო ხელმძღვანელი: ფილოლოგიის მეცნიერებათა კანდიდატი,  
პროფესორი თინათინ მარგალიტაძე

სადისერტაციო ნაშრომის კონსულტანტი: ლიდსის უნივერსიტეტის  
ასოცირებული პროფესორი  
სერგეი შაროვი



უნივერსიტეტის  
გამომცემლობა

თბილისი

2015

## შინაარსი

შესავალი .....	4
I თავი. ლიტერატურის მიმოხილვა .....	10
1.1 სასწავლო ლექსიკოგრაფიის განვითარების ისტორია.....	10
2. კორპუსული ლინგვისტიკა და კორპუსები .....	14
2.1 კორპუსული ლინგვისტიკა .....	14
2.2 კორპუსი.....	17
2.3 ენათა კორპუსები.....	20
2.3.1 ქართული ენის კორპუსები.....	21
2.3.2 ქართული დიალექტური კორპუსი.....	22
2.3.3. ქართული ენის კორპუსი სიტყვათშეხამებითი მოდელების პროგრამაში.....	23
2.3.4 ქართული ენის პარალელური კორპუსი.....	25
3. ვები როგორც კორპუსი.....	27
3.1 ზოგადი მიმოხილვა .....	27
3.2. ვებკორპუსები .....	28
II თავი. ქართული ენის ვებკორპუსი - KaWaC.....	33
1.1 ზოგადი მიმოხილვა .....	33
1.2 ქართული ენის ვებკორპუსი - KaWaC.....	33
1.3 ვებგვერდების ლინგვისტური დამუშავება.....	36
1.4 რეპრეზენტაციულობა ქართული ენის ვებკორპუსში.....	40
2. უანრული კლასიფიკაცია ქართული ენის ვებკორპუსში.....	42
3. მულტიტექსტური მორფოსინტაქსური პარამეტრები.....	47
3.1 მეტყველების ნაწილების მარკერების სისტემა ქართული ენისათვის.....	48
3.2 მორფოლოგიური ანალიზი.....	56
4. სიხშირული სია (Frequency List).....	62

III თავი. ქართული ენის სიხშირული ლექსიკის განსაზღვრა .....	64
1.1 ზოგადი მიმოხილვა .....	64
1.2 კორპუსზე დაფუძნებული სიხშირული ლექსიკონები.....	64
2. ქართული ენის ვებკორპუსის ლემატიზება და ქართული ენის სიხშირული ლექსიკის განსაზღვრა.....	70
2.1 ქართულ-ევროპული ტიპის სასწავლო ლექსიკონის სიხშირული სიტყვანი.....	72
2.2 ქართული სიხშირული ლექსიკის სიტყვანის შედარება „ქართულ ლექსიკონთან“ .....	78
IV თავი, ქართული ენის ვებკორპუსი –KaWaC, როგორც წყარო ქართულ-ევროპული ტიპის სასწავლო ლექსიკონის სიტყვა-სტატიის შესადგენად.....	83
1.1 ზოგადი მიმოხილვა .....	83
1.2 სიხშირული პრინციპით გაანალიზებული „ენ-გრამები“ (n-grams) ქართული ენის ვებკორპუსის მიხედვით .....	83
2. ქართული ენის ვებკორპუსი, როგორც ქართულ-ევროპული ტიპის სასწავლო ლექსიკონის წყარო.....	93
2.1 ქართული ენის ვებკორპუსის მიხედვით შედგენილი სალექსიკონო სიტყვა-სტატიები და მათი შედარება ქართული ენის განმარტებით ლექსიკონთან და „ქართულ ლექსიკონთან“ .....	94
დასკვნა.....	156
გამოყენებული ლიტერატურა.....	162

## შესავალი

სადისერტაციო ნაშრომში წარმოვადგენთ ქართულ-ევროპული ტიპის ახალი სასწავლო ლექსიკონების შედგენის საკითხებს კორპუსული მეთოდოლოგიის გამოყენებით.

XX საუკუნის 80-იანი წლებიდან მოყოლებული, ელექტრონული რესურსებისა და ენათა ეროვნული კორპუსების შექმნამ მთლიანად შეცვალა დამოკიდებულება როგორც ენების კვლევისადმი, ისე ლექსიკონების შედგენისადმი. თანამედროვე მიდგომებით ენის კვლევა წარმოუდგენელია კორპუსული მონაცემების გარეშე. კორპუსზე დაფუძნებული და კორპუსით წარმართული მეთოდოლოგია განსაკუთრებით აქტუალურია ლექსიკოგრაფიული საქმიანობისათვის – ერთენოვანი თუ ორენოვანი ლექსიკონების, განსაკუთრებით კი სასწავლო ტიპის ლექსიკონების შედგენისათვის.

სადოქტორო კვლევის მნიშვნელოვანი ნაწილი შესრულდა ღიდ ბრიტანეთში, ლიდსის უნივერსიტეტში, ამავე უნივერსიტეტის პროფესორის სერგეი შაროვის მონაწილეობით. ლიდსის უნივერსიტეტში კვლევა ორ ეტაპად განხორციელდა: პირველ ეტაპი მოიცავდა 3-თვიან პერიოდს (2014 წლის მარტი-მაისი). სადოქტორო კვლევის პირველი ეტაპი დაფინანსდა შოთა რუსთაველის ეროვნული სამეცნიერო ფონდის მიერ გამოცხადებული დოქტორანტურის საგანმანათლებლო პროგრამების საგრანტო კონკურსის ფარგლებში, ხოლო კვლევის მეორე ეტაპი განხორციელდა 2015 წლის მარტში.

**სადოქტორო კვლევის მიზანი** იყო ქართულ-ევროპული ტიპის ახალი სასწავლო ლექსიკონების შედგენის თეორიული საკითხების დამუშავება კორპუსული მეთოდოლოგიის გამოყენებით. კერძოდ, მიზნად დავისახეთ, ერთი მხრივ, დაგვედგინა ლექსიკონების მაკროსტრუქტურა, ანუ ქართულ-ევროპული ტიპის სასწავლო ლექსიკონების ქართული ნაწილის სიტყვანი სიხშირული პრინციპის მიხედვით; მეორე მხრივ კი, განგვესაზღვრა საკუთრივ სალექსიკონო სიტყვა-სტატიის შედგენის ეფექტურობა კორპუსული მეთოდოლოგიის გამოყენებით, ქართული სიტყვების პოლისემიური მნიშვნელობების სიხშირული პრინციპით გამოვლენა, ყველაზე ხშირად ხმარებული შესიტყვებების, იდიომების დადგენა და სხვა საკითხები.

აღნიშნული მიზნების მისაღწევად საჭირო გახდა ქართული ენის ვრცელი ვებკორპუსის შექმნა. როგორც ზემოთ აღინიშნა, ქართული ენის ვებკორპუსი

შეიქმნა ლიდსის უნივერსიტეტში და სადოქტორო კვლევის მიერ მიზნად დასახული ამოცანები სწორედ ჩვენ მიერ შექმნილი ვებკორპუსის ბაზაზე განხორციელდა: განისაზღვრა 10 000 ყველაზე ხშირად ხმარებული ქართული სიტყვა, რომელიც გახდება ქართულ-ევროპული სასწავლო ლექსიკონების მაკროსტრუქტურის შექმნის საფუძველი; ასევე, გამოვიკვლიეთ ქართული ენის ვებკორპუსის ეფექტურობა ქართულ-ევროპული ტიპის სასწავლო ლექსიკონების ქართული ნაწილის სიტყვა-სტატიების შედგენისას.

უნდა აღინიშნოს, რომ კორპუსული მეთოდოლოგიის გამოყენება სიახლეს წარმოადგენს ქართულ ლექსიკოგრაფიაში. სადოქტორო ნაშრომის **ძირითადი სიახლეებია:**

**1. ქართული ენის ვებკორპუსი – KaWaC.** პირველი და ამ ეტაპზე ერთადერთი ვებკორპუსი ქართული ენისათვის, რომელიც შეიქმნა ვებკორპუსების უახლოესი მეთოდებისა და სტანდარტების შესაბამისად. ქართული ენის ვებკორპუსი ასახავს თანამედროვე, ცოცხალ ქართულ ენას. კერძოდ, ენის განვითარების ბოლო 10-15 წელიწადს. ქართული ენის ვებკორპუსი მოიცავს 150 მილიონზე მეტ სიტყვას და განთავსებულია ლიდსის უნივერსიტეტის ვებგვერდზე (<http://corpus.leeds.ac.uk/internet.html>);

**2. ქართული ენის სიხშირული ლექსიკა.** ქართული ენის ვებკორპუსის მიხედვით განისაზღვრა ყველაზე ხშირად ხმარებული 10 000 სიტყვა, რომელიც წინამდებარე სადისერტაციო ნაშრომს ელექტრონულად აქვს დართული ორი ვერსიის სახით: პირველ ვერსიაში აღნიშნული სიტყვანი დალაგებულია სიხშირული ზრდის მიხედვით, ხოლო მეორე ვერსია ანბანთრიგზეა დალაგებული და მითითებული აქვს შესაბამისი სიხშირული ინდექსი.

გარდა ამისა, სადოქტორო ნაშრომის **სიახლეებია:**

**ა) ფუნქციური მნიშვნელობის მქონე 300 ქართული სიტყვა (Function Words),** რაც წინამდებარე სადოქტორო ნაშრომს ელექტრონული სახით აქვს დართული.

ფუნქციური სიტყვები ამა თუ იმ ენის ყველაზე ხშირად ხმარებულ ლექსიკას წარმოადგენს. ფუნქციური სიტყვების განსაზღვრა უშუალოდ ვებგვერდებიდან ტექსტური მასალის ჩამოტვირთვამდე ძალიან მნიშვნელოვან ეტაპს წარმოადგენდა კორპუსის შექმნის პროცესში, რათა გამოგვერიცხა არაქართული გვერდები, ანიმაციები და ლინგვისტურად სხვა არარელევანტური ინფორმაცია.

**ბ) ვებკორპუსის დაბალანსება / ჟანრული კლასიფიკაცია.** ქართული ენის ვებკორპუსში ტექსტური მასალის დაბალანსება განხორციელდა, ე.წ. „ფუნქციური ჟანრული განზომილებების“ (Functional Genre Dimensions) მიხედვით, რაც ამ ნაშრომის მნიშვნელოვან სიახლეს წარმოადგენს.

**გ) ქართული ენისათვის მეტყველების მარკერების სისტემის, ე.წ. ტაგსეტების (tagset) შექმნა „მულტიტექსტური მორფოსინტაქსური პარამეტრების“ (MULTEXT-East Morphosyntactic Specifications - MTE) მეოთხე ვერსიის მიხედვით.** აღნიშნული ვერსია, რომელიც წარმოადგენს მრავალენოვან მონაცემთა ბაზას ენის ლინგვისტური დამუშავებისათვის და მოიცავს 16 ცენტრალურ და აღმოსავლეთ ევროპულ ენებს (<http://nl.ijs.si/ME/V4/msd/html/index.html>), პირველადაა გამოყენებული ქართული ენისათვის. ქართული ენის მეტყველების მარკერების სისტემა სრული სახით წარმოდგენილია წინამდებარე სადოკუმენტო ნაშრომში.

**დ) ქართული ენის სიხშირული „ენ-გრამები“ ქართული ენის ვებკორპუსის მიხედვით.** სადოკუმენტო ნაშრომის ერთ-ერთი ძირითად სიახლეს წარმოადგენს სიხშირული „ენ-გრამები“, რაც გულისხმობს მომიჯნავე / მოსაზღვრე / მეზობელი ფონემების, სიტყვების, სიტყვათა წყვილების ერთად რეალიზების სიხშირულ გაანალიზებას. სადოკუმენტო კვლევაში სიხშირულად გავანალიზეთ და წარმოვადგინეთ ორკომპონენტიანი, სამკომპონენტიანი, ოთხკომპონენტიანი და ხუთკომპონენტიანი სიხშირული „ენ-გრამები“ ქართული ენის ვებკორპუსის მიხედვით, რაც წინამდებარე სადოკუმენტო ნაშრომს ელექტრონული სახით აქვს დართული.

**ე) ქართული ენის ვებკორპუსის მიხედვით შედგენილი სალექსიკონო სიტყვა-სტატიები.** სადისერტაციო ნაშრომში საილუსტრაციოდ წარმოდგენილია ქართული ენის ვებკორპუსის მიხედვით შედგენილი სალექსიკონო სიტყვა-სტატიების ნაწილი, რომლებიც შედარებულია ქართული ენის განმარტებითი ლექსიკონისა და „ქართული ლექსიკონის“ შესაბამის სიტყვა-სტატიებთან. შედარების საფუძველზე მიღებული შედეგების გაანალიზებამ აჩვენა, რომ ქართული ენის ვებკორპუსის მიხედვით შედგენილი სალექსიკონო სიტყვა-სტატიები გაცილებით მოცულობითია, როგორც საილუსტრაციო მასალის მრავალფეროვნების, ასევე შესიტყვებების თვალსაზრისით. გარდა ამისა, ქართული ენის ვებკორპუსის მიხედვით შედგენილ სიტყვა-სტატიებში ასახულია

ქართული სიტყვების ახალი მნიშვნელობები, რაც ზემოსხენებულ ლექსიკონებში არაა დადასტურებული და ვფიქრობთ, რომ ამ მნიშვნელობების შეტანა ლექსიკონებში და განსაკუთრებით კი, სასწავლო ლექსიკონებში აუცილებელია.

**დისერტაციის თეორიული და მეთოდოლოგიური საფუძველია** სასწავლო /სიხშირული ლექსიკონების, ვებკორპუსების შესახებ არსებული ლიტერატურა, უახლესი კვლევები და სამეცნიერო ნაშრომები. სადოქტორო კვლევის თითოეულ ეტაპზე დავეყრდნით მოცემული საკითხის შესახებ არსებულ უახლეს მეთოდოლოგიას. 1. უპირველეს ყოვლისა, სადოქტორო დისერტაციის – „ქართულ-ევროპული ტიპის ახალი სასწავლო ლექსიკონების შედგენის საკითხები კორპუსული მეთოდოლოგიის გამოყენებით“ თემატიკიდან გამომდინარე გავითვალისწინეთ სასწავლო ლექსიკოგრაფიის ქანრული თავისებურებები, რომლებიც ლექსიკური ერთეულების სიხშირულ შერჩევას უკავშირდება. გარდა სასწავლო ლექსიკოგრაფიის მეთოდოლოგიისა, სადოქტორო დისერტაციის თეორიული და მეთოდოლოგიური საფუძველია უახლესი კორპუსული მიდგომები. ჩვენს კვლევაში გამოვიყენეთ „სიხშირული ლექსიკონების“ გამოცდილება და ამ კუთხით არსებული კვლევები და გაგაანალიზეთ არსებული სიხშირული ლექსიკონები: რუსული, ფრანგული, ჩეხური და სხვა ენების სიხშირული ლექსიკონები. სიხშირული ლექსიკონების ერთ-ერთი ძირითადი პრინციპია ამა თუ იმ ენის სიხშირული ლექსიკის განსაზღვრა ვებკორპუსების ბაზაზე.

2. ქართული ენის ვებკორპუსის შექმნა რამდენიმე ეტაპად წარიმართა. შესაბამისად, კვლევის ცალკეულ ეტაპზე გამოვიყენეთ შესაბამისი მეთოდოლოგიები, კერძოდ:

- 1) ვებმასალის ჩამოტვირთვის პროცესში გამოვიყენეთ სპეციალური კორპუსული პროგრამა - “wget” (World Wide Web get);
- 2) ვებგვერდიდან ჩამოტვირთული მასალა ლინგვისტურად დამუშავდა ასევე სპეციალური კორპუსული ინსტრუმენტის “BootCatTools”-ის საშუალებით. აღსანიშნავია, რომ ვებრესურსების ლინგვისტური დამუშავებისათვის გამოვიყენეთ, ე.წ. „კლინივალის“ (CleanEval: cleaning webpages) გამოცდილება (Baroni et al 2008);
- 3) ქართული ენის ვებკორპუსის დაბალანსებისათვის და ქანრული კლასიფიკაციისათვის გამოვიყენეთ მელერის, შაროვისა და სანტინის

(Mehler et al 2010) ვების ქანრული კლასიფიკაციები – „ფუნქციური ქანრული განზომილებები“ (Functional Genre Dimensions);

- 4) ქართული ენის ვებკორპუსის მორფოლოგიური ანოტირების პლატფორმად გამოყენებულია „მულტიტექსტური მორფოსინტაქსური პარამეტრების“ (MULTEXT-East Morphosyntactic Specifications - MTE) მეოთხე ვერსია;
- 5) ქართული ენის ვებკორპუსის ლემატიზების პროცესში გამოვიყენეთ CST ლემატიზების ინსტრუმენტი (CST Lemmatiser), რომელიც შემუშავებულია კოპენჰაგენის უნივერსიტეტის ენის ტექნოლოგიის ცენტრში.

ამგვარად, წინამდებარე დისერტაციის თეორიულ და მეთოდოლოგიურ საფუძველს წარმოადგენს სასწავლო და სიხშირული ლექსიკონებისა და ვებკორპუსების შესახებ არსებული უახლესი კვლევები და სამეცნიერო ნაშრომები. სადოქტორო კვლევა ეხმიანება გამოყენებითი ენათმეცნიერების დღევანდელ მდგომარეობას, კერძოდ, ლექსიკოგრაფიისა და კორპუსული ლინგვისტიკის სინთეზს.

წინამდებარე სადისერტაციო ნაშრომი „ქართულ-ევროპული ტიპის ახალი სასწავლო ლექსიკონების შედგენის საკითხები კორპუსული მეთოდოლოგიის გამოყენებით“ შედგება 4 თავისაგან, რაც თავის მხრივ, 30 ქვეთავად არის დაყოფილი ნაშრომში.

**I თავი. ლიტერატურის მიმოხილვა.** სადოქტორო ნაშრომის პირველი თავი ეხება ლიტერატურის მიმოხილვას, ამ თავში აღწერილია სასწავლო ლექსიკოგრაფიის განვითარების ისტორია და პირველი სასწავლო ლექსიკონები, კორპუსული ლინგვისტიკის განვითარების ისტორია, კორპუსის რაობა, კორპუსების კლასიფიკაცია და ენათა ისეთი კორპუსები, როგორებიცაა, ბრაუნის კორპუსი, ბრიტანული ნაციონალური კორპუსი. ასევე, ქართული ენისათვის შექმნილი კორპუსები: თანამედროვე ქართული ენის კორპუსი (GEKKO), ქართული დიალექტური კორპუსი, ქართული ენის კორპუსი სიტყვაშეხამებითი მოდელის პროგრამაში და ქართული ენის პარალელური კორპუსი. ამავე თავში განხილულია კორპუსების ახალ მიმართულება – WaC, „ვები როგორც კორპუსი“ (Web as a Corpus).

**II თავი. ქართული ენის ვებკორპუსი - KaWaC.** სადოქტორო ნაშრომის მეორე თავი ეხება ქართული ენის ვებკორპუსის – KaWaC-ის შექმნის ეტაპებსა და ამ მიზნით ლიდსის უნივერსიტეტში შესრულებულ სამუშაოებს, მაგალითად,



როგორცაა ვებგვერდების ლინგვისტური დამუშავება, კერძოდ, ვებგვერდების ანოტირება და გასუფთავება; ვებკორპუსის დაბალანსება, მეტყველების ნაწილების მარკერების სისტემების შექმნა, მორფოლოგიური ანოტირება და ა.შ.

**III თავი. ქართული ენის სიხშირული ლექსიკის განსაზღვრა.** აღნიშნული თავი ეხება სიხშირულ ლექსიკონებსა და ქართული ენის სიხშირული ლექსიკის განსაზღვრისათვის ჩატარებულ სამუშაოებს, კერძოდ, ქართული ენის ვებკორპუსის ლემატიზებას.

**IV თავი. ქართული ენის ვებკორპუსი - KaWaC, როგორც წყარო ქართულ-ვეროპული ტიპის სასწავლო ლექსიკონისათვის.** მეოთხე თავში განხილულია ქართული ენის ვებკორპუსი და მისი ტექსტური მასალა და წარმოდგენილია ქართული ენის ვებკორპუსის მიხედვით შედგენილი სალექსიკონო სიტყვა-სტატიები. ნაშრომის შედეგები დასკვნაშია განზოგადებული. ნაშრომს დართული აქვს გამოყენებული ლიტერატურის ნუსხა და დამატებითი მასალა.

წინამდებარე სადისერტაციო ნაშრომი შეიძლება გამოყენებულ იქნეს როგორც ქართულ-ვეროპული ტიპის ახალი სასწავლო ლექსიკონების, ასევე სიხშირული ლექსიკონების შედგენისათვის და ქართული ენის დიდი განმარტებითი ლექსიკონის შევსებისათვის. ნაშრომში წარმოდგენილი უახლესი კორპუსული მეთოდები და მიდგომები შეიძლება გამოყენებულ იქნეს სხვა ქართველური ენების კორპუსების, კორპუსული რესურსების შექმნისათვის და მსგავსი კორპუსული კვლევის განსახორციელებლად.

## I თავი. ლიტერატურის მიმოხილვა

### 1.1 სასწავლო ლექსიკოგრაფიის განვითარების ისტორია

XX საუკუნის პირველი ნახევარიდან ინგლისური ენის სწავლების მოთხოვნის გაზრდამ მნიშვნელოვანად შეუწყო ხელი ლექსიკოგრაფიის ახალი ენის - სასწავლო ლექსიკოგრაფიის განვითარებას. ინგლისური ენის შემსწავლელთათვის აუცილებელი გახდა, ე.წ. „მინიმალური ლექსიკის“ (Controlled Vocabulary -VC) განსაზღვრა, ვინაიდან ინგლისური ენის განმარტებით ლექსიკონებში მოცემული სალექსიკონო სიტყვა-სტატიის განმარტებები რთულად გასაგები იყო ინგლისური ენის შესწავლით დაინტერესებული უცხოელებისათვის და საჭირო იყო ინგლისური ენის „ლექსიკური მინიმუმის“ განსაზღვრა, რომელიც მათ ენის შესწავლას გაუადვილებდა.

სასწავლო ლექსიკოგრაფიის განვითარება დაკავშირებულია ისეთი ენათმეცნიერების სახელებთან, როგორებიცაა: ალბერტ ჰორნი, მაიკლ ვესტი და ჰაროლდ პალმერი. სწორედ მათი ინიციატივით შეიქმნა მოძრაობა, ე.წ. „მინიმალური ლექსიკის განსაზღვრისათვის“ (The Vocabulary Control Movement) და მათ სავსებით სამართლიანად, ბრიტანეთის გამოყენებითი ენათმეცნიერების მამებსაც კი უწოდებენ (“the fathers of British applied linguistics”).

განსაკუთრებით საინტერესოა ის ფაქტი, რომ პირველი სასწავლო ლექსიკონები არა ინგლისში, არამედ იაპონიაში, ტოკიოს ინსტიტუტის ინგლისური ენის სწავლების ცენტრში (Tokyo Institute for Research in English Teaching (IRET) მომზადდა და გამოიცა. აღნიშნული ცენტრი დაარსდა 1923 წლის მაისში ჰაროლდ პალმერის მიერ. ჰაროლდ პალმერი დაინტერესებული იყო იმ მინიმალური ლექსიკის განსაზღვრით, რაც ყველაზე ხშირად გამოიყენებოდა ყოველდღიურ მეტყველებაში და აუცილებელი იყო მინიმალური კომუნიკაციის დასამყარებლად. პალმერი და მისი თანამოაზრეები ეფუძნებოდნენ მათ უშუალოდ საკუთარ გამოცდილებას სწავლების პროცესში (Palmer 1920) და მათი მეთოდოლოგია, ე.წ. „სუბიექტურ“ (subjective approach) ხასიათს ატარებდა, განსხვავებით მათი ამერიკელი კოლეგებისგან, რომელთა შორის უნდა აღინიშნოს ედუარდ თორნდაიკი კოლუმბიის უნივერსიტეტიდან. იგი, ე.წ. „ობიექტურ“ (objective approach) მიდგომას უჭერდა მხარს. „ობიექტური“ მიდგომა გულისხმობდა დაბალანსებული ტექსტების ბაზაზე სისწორული პრინციპით

სიტყვანის შერჩევას (Thorndike 1921). 1934 წელს, ნიუ-იორკში ჩატარდა კარნეგის კონფერენცია ინგლისურის, როგორც უცხო ენის სწავლებისათვის ლექსიკის შერჩევის თაობაზე (the Carnegie Conference on Vocabulary Selection for the Teaching of English as a Foreign Language), სადაც კონფერენციის მონაწილეებმა მიიღეს გადაწყვეტილება შეეჯერებინათ „ობიექტური“ და „სუბიექტური“ მიდგომები ინგლისური ენის საბაზო ლექსიკის განსაზღვრისათვის (Palmer, West, Faucett, 1936).

1938 წელს, ჰაროლდ პალმერმა ალბერტ ჰორნბისთან ერთად შეიმუშავა 1000 მინიმალური სიტყვა (Thousand-Word English), რომელიც მიზნად ისახავდა ინგლისური ენის შემსწავლელთათვის ენის შესწავლის გაადვილებას (Palmer & Hornby 1938). მაიკლ ვესტმა კი უფრო ვრცელი ლექსიკა წარმოადგინა, დაახლოებით 1,500 ერთეულის შემცველი „მინიმალური ლექსიკა“ ინგლისური ენისა. სასწავლო ლექსიკოგრაფიის ისტორიაში, პირველ სასწავლო ლექსიკონს წარმოადგენს „ახალი მეთოდების მიხედვით შედგენილი ინგლისური ენის ლექსიკონი“ (A New Method English Dictionary), რომელიც ეკუთვნის მაიკლ ვესტსა და ჯეიმს ენდიკოტს (West & Endicott, 1935). აღნიშნული ლექსიკონი გამოიცა 1935 წელს და მოიცავდა 23,898 სალექსიკონი სიტყვა-სტატიას, რომლებიც მაიკლ ვესტის მიერ განსაზღვრული 1,490 მინიმალური ლექსიკით /სიტყვით იყო განმარტებული.

ამავე პერიოდში, გამოიცა არაერთი სასწავლო ლექსიკონი ინგლისური ენის „მინიმალური ლექსიკის“ მიხედვით, მათ შორის აღნიშვნის ღირსია პალმერის (Palmer 1938) „ინგლისური სიტყვების გრამატიკა“ (A Grammar of English Words), ჰორნბის და იშიკავას (Hornby & Ishikawa 1940) „ინგლისურ-იაპონური ლექსიკონი დამწყებთათვის“ (A Beginner's English-Japanese Dictionary). განსაკუთრებით საინტერესოა ოგდენის (Ogden 1935) „ზოგადი ინგლისური ენის ლექსიკონი“ (The General Basic English Dictionary), ოგდენის ლექსიკონი მოიცავდა 20,000 სალექსიკონო ერთეულს, რომლებიც 850 საბაზო ინგლისური სიტყვით იყო განმარტებული.

უფრო მოგვიანებით, 1953 წელს, ვესტმა (West 1953) წარმოადგინა „ზოგადი გამოყენების ლექსიკა“ (General Service List), რომელიც მოიცავდა როგორც ცალკეული სიტყვების, ასევე შესიტყვებებისა და იდიომების სისშირულ

ინდექსებს (frequency ratings), რამაც ცალსახად მნიშვნელოვანი გავლენა იქონია სასწავლო ლექსიკოგრაფიის შემდგომ განვითარებაზე.

ინგლისური ენის სასწავლო ლექსიკონების შემდგომი განვითარება აისახა, ე.წ. „დიდ ხუთეულში“ (“Big-Five”), რომლებიც წარმოადგენენ ინგლისური ენის ყველაზე მნიშვნელოვან და ძირითად სასწავლო ლექსიკონებს (Tarp 2008). ესენია:

- 1) **ოქსფორდის სასწავლო ლექსიკონი** (Oxford Advanced Learner’s Dictionary); აღნიშნული ლექსიკონი პირველად გამოიცა 1948 წელს, ხოლო ბოლო, მე-9 გამოცემა, 2015 წელს;  
(<http://www.oxfordlearnersdictionaries.com/definition/english/>);
- 2) **ლონგმენის თანამედროვე ინგლისური ენის ლექსიკონი** (The Longman Dictionary of Contemporary English (LDOCE)); აღნიშნული ლექსიკონი პირველი პირველად გამოიცა 1978 წელს, ხოლო ბოლო, მე-6 გამოცემა, 2014 წელს  
(<http://global.longmandictionaries.com/>);
- 3) **კოლინზ ქობილდის ინგლისური ენის ლექსიკონი** (The Collins COBUILD English Language Dictionary); ლექსიკონი პირველად გამოიცა გამოვიდა 1987 წელს, ხოლო ბოლო, მე-6 გამოცემა, 2009 წელს;  
(<http://www.collins.co.uk/category/English+Language+Teaching/COBUILD+Reference>);
- 4) **კემბრიჯის ინგლისური ენის საერთაშორისო ლექსიკონი** (The Cambridge International Dictionary of English); ლექსიკონის პირველი გამოცემა გამოვიდა 1995 წელს, ხოლო ბოლო მე-3 გამოცემა 2013 წელს კემბრიჯის უნივერსიტეტის გამომცემლობის მიერ;  
(<http://dictionary.cambridge.org/dictionary/british/>);
- 5) **“მაკმილანის ინგლისური ენის ლექსიკონი“** (The Macmillan English Dictionary); მაკმილანის ლექსიკონი პირველად გამოიცა 2002, ხოლო მეორე, ბოლო გამოცემა, 2007 წელს. 2012 წლის გადაწყვეტილებით მაკმილანის ლექსიკონი არ გამოიცემა ბეჭდური სახით, ხელმისაწვდომია მხოლოდ ონლაინ და ელექტრონული ვერსია (<http://www.macmillandictionary.com/>).

ამგვარად, XX საუკუნის პირველი ნახევრიდან იწყება მუშაობა პირველ სასწავლო ლექსიკონებზე და ვითარდება ლექსიკოგრაფიის ახალი ეტაპი – სასწავლო ლექსიკოგრაფია, რომელსაც მისთვის დამახასიათებელი პრინციპები ჩამოუყალიბდა, რაც სალექსიკონო სიტყვა-სტატიების სისშირული პრინციპით შერჩევას უკავშირდება. კომპიუტერული ტექნოლოგიების განვითარების

კვლადაკვალ დაიხვეწა სასწავლო ლექსიკონების შედგენის მეთოდოლოგია და სასწავლო ლექსიკოგრაფია განვითარების სხვა ეტაპზე გადავიდა. კორპუსისა და კომპიუტერული ლინგვისტიკის განვითარების კვლადაკვალ იქმნება მრავლმილიონიანი და მილიარდსიტყვიანი ენათა კორპუსები, რომლის ბაზაზე შექმნილ ლექსიკონებს „სიხშირულ ლექსიკონებს“ (Frequency Dictionaries) „კორპუსზე დაფუძნებულ“ (corpus-based) და „კორპუსით წარმართულ“ (corpus-driven) ლექსიკონებს უწოდებენ, ხოლო კორპუსების გამოყენებას ლექსიკოგრაფიულ საქმიანობაში კი „კორპუსზე დაფუძნებულ ლექსიკოგრაფიას“. წინამდებარე სადოქტორო ნაშრომი ეფუძნება, როგორც სასწავლო ლექსიკოგრაფიის, ასევე კორპუსულ მეთოდოლოგიას, ე.წ. „სიხშირული ლექსიკონების“ პრინციპებს, რაზეც უფრო დეტალურად, ქვემოთ III თავში ვისაუბრებთ.

## 2. კორპუსული ლინგვისტიკა და კორპუსები

### 2.1 კორპუსული ლინგვისტიკა

კორპუსული ლინგვისტიკა არის გამოყენებითი ენათმეცნიერების ახალი დარგი და გულისხმობს ლინგვისტური კვლევის ახალი, კორპუსული მეთოდით განხორციელებას. ტერმინი „კორპუსული ლინგვისტიკა“. გამომდინარეობს ინგლისური “Corpus Linguistics“-იდან, აქვე უნდა აღინიშნოს, რომ ქართულ ენაში ტერმინ „კორპუსული ლინგვისტიკის“ პარალელურად გამოიყენება „კორპუსის ლინგვისტიკა“, აგრეთვე „კორპუსლინგვისტიკა“, რომელიც ფრანკფურტის სკოლამ შემოგვთავაზა. კორპუსული ლინგვისტიკა არის ცოცხალი და ბუნებრივი ენის შესწავლა და კვლევა ვრცელი მოცულობის კორპუსების საფუძველზე. ჩვეულებრივ, კორპუსული კვლევა ხორციელდება კომპიუტერის საშუალებით, ამა თუ იმ კვლევისათვის სპეციალურად შექმნილი ინსტრუმენტის დახმარებით და ითვალისწინებს შესაბამისი ლინგვისტური მოვლენის სიხშირული პრინციპით შესწავლას და ა.შ. როგორც ვიცით, ენის შესწავლის პროცესში ტექსტებიდან ცოცხალი მაგალითების გამოყენება არ არის სიახლე ლინგვისტიკის ისტორიაში. თუმცა, კორპუსული ლინგვისტიკა განვითარდა და მას ახალი სიცოცხლე შესძინა კომპიუტერული ტექნიკის განვითარებამ. კომპიუტერმა და ელექტრონულ ფორმატში მოცემულმა ტექსტებმა შესაძლებელი გახადა დიდი მოცულობის მასალის სწრაფად და ადვილად შეგროვება, დამუშავება და გაანალიზება. ამიტომ, გასაკვირი არ უნდა იყოს, რომ კორპუსულმა ლინგვისტიკამ მიიღო თავისი თანამედროვე ფორმა მხოლოდ კომპიუტერული რევოლუციის შემდგომ, 1980-იან წლებში.

აღამ კილგარიფი (Kilgarriff 2001) კორპუსული ლექსიკოგრაფიის ისტორიაში გამოყოფს ოთხ პერიოდს:

- 1) **პირველი პერიოდი**, ეს არის პერიოდი კომპიუტერის გამოგონებამდე, როდესაც ლექსიკოგრაფები აგროვებდნენ / აბარებდნენ ციტატებს, მაგალითად, როგორც ეს დიდი ოქსფორდის 20-ტომეული განმარტებითი ლექსიკონის (Oxford English Dictionary- OED) მაგალითზე ხდებოდა;
- 2) **მეორე პერიოდი** კი არის ქობილდის ლექსიკონის შედგენის პერიოდი ადრეულ 1980-იან წლებში, როდესაც ლექსიკოგრაფები აგროვებდნენ ყველა სახის მონაცემებს კორპუსიდან და იყენებენ შედარებით მცირე მოცულობის კორპუსულ მონაცემებს (Sinclair 1987);

3) **მესამე პერიოდი** არის გვიანი 1990-იანი წლები, როდესაც კორპუსი გახდა იმდენად ვრცელი, რომ კორპუსული მასალის დამუშავებისა და გამოყენებისათვის საჭირო გახდა გარკვეული კომპიუტერული ინსტრუმენტი მონაცემების შესაბამისად წარმოდგენისათვის.

4) **მეთხე** პერიოდი არის უკვე თანამედროვე კომპიუტერული პროგრამებითა და ინსტრუმენტებით აღჭურვილი კორპუსული ლინგვისტიკის პერიოდი.

ზოგადად, კორპუსული ლინგვისტიკა ორ პერიოდად იყოფა: **ადრეული კორპუსული ლინგვისტიკა**, ასევე ცნობილია როგორც ხომსკიმდელი კორპუსული ლინგვისტიკა და **თანამედროვე კორპუსული ლინგვისტიკა**. კორპუსული ლინგვისტიკის ადრეული მაგალითები თარიღდება გვიანი XIX საუკუნით და გერმანიაში იღებს სათავეს. 1897 წელს გერმანელმა ლინგვისტმა ჯ. კადინგმა (Käding 1897) გამოიყენა საკმაოდ ვრცელი მოცულობის კორპუსი, რომელიც დაახლოებით 11 მილიონ სიტყვას მოიცავდა (McEnery & Wilson 1996). ასეთი ვრცელი მოცულობის კორპუსის გამოყენება, რომელიც თანამედროვე კორპუსის მოცულობას შეესაბამება, იმ დროს რევოლუციური მოვლენა იყო. იმ ლინგვისტთა შორის, რომლებმაც გამოიყენეს ენის კორპუსები და კორპუსული მეთოდოლოგია როგორც ენის სწავლებისათვის ასევე, კვლევისათვის XIX საუკუნის პირველ ნახევარში, აღნიშვნის ღირსია ფრანც ბოასი (Boas 1911), ზელინგ ჰარისი (Harris 1951), ჩარლზ ფრაიზი (Fries 1952), ლეონარდ ბლუმფილდი (Bloomfield 1933) და სხვები. ჩარლზ ფრაიზმა დაიწყო კორპუსის გამოყენება უცხო ენის პედაგოგიურ სწავლებაში. ამერიკის შეერთებულ შტატებში პირველი კორპუსი გამოჩნდა 1920-იან წლებში. პირველი, თანამედროვე, დიდი ინგლისური ენის კორპუსს ლექსიკური კვლევებისათვის მიმართა ედუარდ თორნდაიკმა, მან გამოიყენა 4.5 მილიონი სიტყვა თავის „მასწავლებლის სახელმძღვანელო წიგნში“ (Teacher's Word Book), რომელიც 1921 წელს გამოსცა (Thorndike 1921). მეორე დიდი მოცულობის კორპუსი იყო ერნესტ ჰორნის „წერის ძირითადი ლექსიკა“ (A Basic Writing Vocabulary), რომელიც 1926 წელს გამოვიდა და მოიცავდა 10,000 სიტყვას (Horn 1926). უნდა აღინიშნოს, რომ კორპუსის ადრეული კვლევები ბრიტანეთსა და საფრანგეთში მიზნად ისახავდა მხოლოდ სიტყვის გამოყენების სიხშირის დადგენას.

1961 წელს ჰენრი კუჩერამ და ნელსონ ფრანცისმა ბრაუნის უნივერსიტეტიდან დაიწყეს თანამედროვე ამერიკული ინგლისური ენის ბრაუნის კორპუსზე მუშაობა. აღნიშნული კორპუსი ცნობილია როგორც ბრაუნის კორპუსი (Brown Corpus) და

არის პირველი თანამედროვე ელექტრონული კორპუსი. ბრაუნის კორპუსი შედგება 1 მილიონი ამერიკული ინგლისური სიტყვისაგან, რომლებიც სხვადასხვა თემატიკის ტექსტებიდანაა აღებული და დაჯგუფებულია 15 თემატურ კატეგორიად (<http://clu.uni.no/icame/brown/bcm.html>). უნდა აღინიშნოს, რომ კორპუსული ლინგვისტიკის თანამედროვე სტანდარტებისათვის ბრაუნის კორპუსის მოცულობა არის მცირე. თუმცა, ბრაუნის კორპუსი მიჩნეულია როგორც ერთ-ერთი მნიშვნელოვანი ნაშრომი კორპუსული ლინგვისტიკის ისტორიაში.

1950-იან წლებში ამერიკასა და სხვა ქვეყნებში კორპუსი ნაკლებად პოპულარული იყო ხომსიკისა და მისი მიმდევრების გავლენის გამო, რომელთა აზრით კორპუსი იყო არასწორი მიმართულების მიმცემი. ხომსკის აზრით, კორპუსში ასახული იყო ის, რაც ამა თუ იმ ენის მატარებელმა თქვა და არა ის, რაც უნდა ეთქვა. მიუხედავად ამისა, კორპუსული ლინგვისტიკამ კვლავ მოიპოვა პოპულარობა 1970-იანი წლებიდან. პირველად ტერმინი „კორპუსული ლინგვისტიკა“ გამოიყენეს 1980-იან წლებში, ეს არის ტექნოლოგიის განვითარების ახალი ხანა და სწორედ ამ პერიოდიდან იწყება კორპუსული ლინგვისტიკის, როგორც დარგის აღორძინება და განვითარება.

რა განსხვავებაა კორპუსულ ლინგვისტიკასა და „ტრადიციულ“ ლინგვისტიკას შორის? ამ საკითხზე ვრცლად საუბრობს პროფ. მარინე ბერიძე, მისი აზრით: „კორპუსული ლინგვისტიკა ძირითად ყურადღებას მეტყველების შესწავლას უთმობს, „ტრადიციული“ ლინგვისტიკა კი ენას; ერთი არჩევს კვანტიტატიურ მეთოდებს მეორე კვალიტატიურს; პირველი ეყრდნობა კვლევის ემპირიულ, მეორე რაციონალისტურ მეთოდებს; პირველი ტექსტს განიხილავს, როგორც გარკვეულ ფიზიკურ მოცემულობას, მეორე როგორც აბსტრაქციას; პირველი სტატისტიკურ მასალას აანალიზებს, მეორე ლოგიკურ მსჯელობას ამჯობინებს და ემპირიულ მასალასაც ინტუიციისაზე დაყრდნობით არჩევს; ერთი მონაცემთა ბაზას იყენებს, მეორე ხელოვნურ მაგალითებს, ან ტექსტიდან იზოლირებულ სიტყვასმარებებს; ერთი ინდუქციურ მეთოდს არჩევს, მეორე დედუქციურს, ერთი ემპირიულ მასალაზე დაფუძნებულ აღმოჩენებს აღიარებს, მეორე პროცედურებზე, შედარებებზე, შეფასებებზე და ა. შ. დაფუძნებულს. კორპუსული ლინგვისტიკის იდეოლოგიის მთავარი ორიენტირს წარმოადგენს განსაკუთრებული ყურადღება ტექსტისადმი (დისკურსისადმი), ენის რაოდენობრივი (კვანტიტატიური) კომპონენტისადმი, ენის სინქრონიული და დიაქრონიული ვარიაციულობისადმი, განსხვავებული მიდგომა ენობრივი ნორმისა და ენობრივი ფაქტისადმი“ (ბერიძე 2013).



## 2.2 კორპუსი

სიტყვა კორპუსი წარმოდგება ლათინური სიტყვიდან “corpus”, რაც ნიშნავს „სხეულს“, „ტანს“. თანამედროვე ლინგვისტიკაში „კორპუსი“ გამოიყენება ტექსტების ელექტრონული კრებულის აღნიშვნისათვის. მაკენერისა და უილსონის მიხედვით კორპუსი შეიძლება ვუწოდოთ ერთზე მეტი ტექსტის კრებულს (McEnery & Wilson 1996), თუმცა ენათმეცნიერებაში „კორპუსს“ თავისი აუცილებელი პარამეტრი და საზომი გააჩნია. თანამედროვე ენათმეცნიერული გაგებით კორპუსს გააჩნია ოთხი ძირითადი პარამეტრი, მაკენერისა და უილსონის მიხედვით ეს პარამეტრებია:

- 1) შერჩევა და რეპრეზენტაციულობა (“Sampling and Representativeness”);
- 2) სასრულობა (“Finite size”);
- 3) ელექტრონულად წაკითხვადობა (“Machine-readable”);
- 4) სტანდარტულობა (“Standard Reference”).

მსოფლიოს თითქმის ყველა ენისათვის შექმნილია და იქმნება სხვადასხვა ტიპის კორპუსები. ერთი-ერთი კლასიკური და ტრადიციული კორპუსია ბრიტანული ნაციონალური კორპუსი (British National Corpus -BNC), რომელიც არაერთი ენის კორპუსის შექმნის მოდელად არის გამოყენებული, მათ შორის სლავურ ენათა კორპუსებისათვის, რომელთა შორის აღსანიშნავია ჩეხური ენის ნაციონალური კორპუსი (The Czech National Corpus -CNC).

რა აქცევს ტექსტების კრებულს კორპუსად? უპირველეს ყოვლისა, კორპუსისათვის დამახასიათებელია რეპრეზენტაციულობა. მნიშვნელოვანია როგორც ტექსტური მასალის მოცულობა, ასევე პროპორციულობა და დაბალანსება კორპუსის ტექსტური ბაზისა. შეიძლება ითქვას, რომ კორპუსი წარმოადგენს ენის შემცირებულ მოდელს. ამიტომ, კორპუსში მოცემული ტექსტური მასალა უნდა იყოს კარგად დაბალანსებული და მოიცავდეს სხვადასხვა პერიოდის, ჟანრის, სტილის, ავტორის და ა.შ. ტექსტების კრებულს.

კვლევის მიზნებიდან გამომდინარე გამოყოფენ კორპუსის სხვადასხვა ტიპებს (Kennedy 1998). ქვემოთ წარმოდგენილია კორპუსის შემდეგი ტიპები:

- 1) **ზოგადი კორპუსები** (general/reference corpora), მაგალითად, როგორცაა ბრიტანული ნაციონალური კორპუსი (BNC – British National Corpus). ზოგადი კორპუსი მიზნად ისახავს ამა თუ იმ ენის ან მის რომელიმე ვარიაციის

- მოდელის წარმოდგენას, იგი მოიცავს როგორც წერილობით, ასევე ზეპირი მეტყველების ტექსტებს, სხვადასხვა ჟანრსა თუ სტილს და ა.შ.;
- 2) **დარგობრივი/სპეციალიზებული კორპუსები** (specialized corpora), ამა თუ იმ კონკრეტული დარგის კორპუსი, მაგალითად, ბიოლოგიური და ა.შ.;
  - 3) **ისტორიული კორპუსები** (historical corpora), მაგალითად, როგორიცაა ჰელსინკის კორპუსი (Helsinki Corpus, ARCHER <http://www.helsinki.fi/varieng/CoRD/corpora/HelsinkiCorpus/>).  
ისტორიული კორპუსებში წარმოდგენილია ენის განვითარების ადრეული ეტაპები;
  - 4) **რეგიონული კორპუსები** (regional corpora), რეგიონული კორპუსები იმით განსხვავდება ზოგადი ტიპის კორპუსებისაგან, რომ იგი მოიცავს ამა თუ იმ ენის რამდენიმე რეგიონალურ ვარიაციას. რეგიონული კორპუსის ერთ-ერთ მაგალითს წარმოადგენს ახალი ზელანდიური ინგლისურის ველინგტონის კორპუსი – WCNZE (Wellington Corpus of Written New Zealand English) - <http://clu.uni.no/icame/manuals/WELLMAN/INDEX.HTM>;
  - 5) **სასწავლო კორპუსები** (learner corpora VS Native speaker corpora), როგორიცაა, სასწავლო ინგლისურის საერთაშორისო კორპუსი – ICLE (International Corpus of Learner English) - <https://www.uclouvain.be/en-cecl-icle.html>. ამ ტიპის კორპუსები მოიცავს ტექსტებს მოცემული ენის შემსწავლელთათვის, რომელთათვის ეს ენა არ არის მშობლიური ენა;
  - 6) **მრავალენოვანი/პარალელური კორპუსები** (multilingual/parallel corpora) ასახვს, სულ მცირე, ორი სხვადასხვა ენის კორპუსს, ხშირ შემთხვევაში, იდენტურ ტექსტებს, რომლებიც გამოიყენება შედარებითი ანალიზისათვის;
  - 7) **ზეპირი მეტყველების კორპუსები** (spoken corpora), ამ ტიპის კორპუსებში მოცემული მხოლოდ ზეპირი მეტყველები, ასეთი კორპუსია, მაგალითად, LLC – ზეპირი მეტყველების ინგლისურის ლონდონი-ლუნდის კორპუსი (London-Lund Corpus of Spoken English) - <http://clu.uni.no/icame/manuals/LONDLUND/INDEX.HTM>.
  - 8) **წერილობითი კორპუსები** (written corpus), ასეთი ტიპის კორპუსებში მოცემულია მხოლოდ წერილობითი ხასიათი ტექსტები, ასეთი ერთ-ერთი ყველაზე გამორჩეული კორპუსია BAWE – ბრიტანული აკადემიური

ინგლისურის კორპუსი (British Academic Written English Corpus) - <http://www.coventry.ac.uk/bawe>.

ამგვარად, კორპუსი შეიძლება ასახავდეს ამა თუ იმ ენის ზეპირ მეტყველებას მხოლოდ, ან წერილობით ენას, ან ორივეს ერთად, ასეთი ტიპის კორპუსებს, როდესაც მოცემულია როგორც ზეპირი, ასევე წერილობითი მასალა „შერეულ კორპუსებს“ (mixed corpora) უწოდებენ. შერეული ტიპის კორპუსის ერთ-ერთ თვალსაჩინო მაგალითს წარმოადგენს ბრიტანული ნაციონალური კორპუსი, რომელიც მოიცავს როგორც ზეპირი მეტყველების, ასევე წერილობითი ხასიათის ტექსტებს.

ზემოთ წარმოდგენილი კორპუსის კლასიფიკაცია წარმოდგენილია ტექსტის ტიპების მიხედვით. გარდა ამგვარი კლასიფიკაციისა, კორპუსული ლინგვისტიკაში გამოყოფენ კლასიფიკაციას ტექსტური მასალის დამუშავების მიხედვით, მაგალითად, როგორიცაა:

- 1) **ანოტირებული კორპუსები** (annotated corpora), ამ ტიპის კორპუსში განხორციელებულია გარკვეული ლინგვისტური ანალიზი, როგორცაა მორფოლოგიური, სინტაქსური ანალიზი, ლექსიკონი და ა.შ.;
- 2) **ორთოგრაფიული კორპუსები** (orthographic corpora), ამგვარ კორპუსებში მორფოლოგიურ ანოტირებასთან ერთად მოცემულია ამა თუ იმ ფონემის სისწორული სტატისტიკა და ა.შ. ორთოგრაფიული ტიპის კორპუსის ერთ-ერთი მაგალითია კოლფის (Colfis Orthographic Corpus) ორთოგრაფიული კორპუსი (Laudanna, Thorton, Brown, Burani, & Marconi, 1995; Bertinetto et al., 2005), <http://linguistica.sns.it/CoLFIS/Home.htm>.

## 2.3 ენათა კორპუსები

წინამდებარე ქვეთავში მიმოვიხილავთ იმ კორპუსებს, რომლებიც დღევანდელი კორპუსული ლინგვისტიკის გადმოსახედიდან წარმოადგენს, ე.წ. „კლასიკურ“ კორპუსებს, რომლებიც გამოყენებულია და გამოიყენება მრავალი ენის კორპუსების შექმნისათვის. ერთ-ერთ ასეთ კორპუსს წარმოადგენს ბრაუნის კორპუსი - თანამედროვე ამერიკული ინგლისური ენის ბრაუნის სტანდარტული კორპუსი (Brown Corpus (the Brown Standard Corpus of Present-Day American English)). აღნიშნული კორპუსი 1963-64 წლებში შეიქმნა ბრაუნის უნივერსიტეტში ჰენრი კუჩერასა და ნელსონ ფრანციზის მიერ და კორპუსზე მუშაობა 1967 წელს დასრულდა. ბრაუნის კორპუსი შედგება დაახლოებით ინგლისური ენის 500 ტექსტისაგან, რომელიც 15 უანრად არის დაყოფილი და ჯამში შეადგენს 1 მილიონ სიტყვას. ბრაუნის კორპუსმა მნიშვნელოვანი გავლენა იქონია სხვა კორპუსების შექმნაში, რომელთა შორის უნდა აღინიშნოს LOB-ის (Lancaster, Oslo, Bergen) კორპუსი, რომელიც ლანკასტერის უნივერსიტეტში შეიქმნა 1978 წელს. ბრაუნის კორპუსი თანამედროვე კორპუსებთან შედარებით არის მცირე მოცულობის, მაგრამ ყოველთვის წარმოდგენილია ცნობილ და მნიშვნელოვან კორპუსთა შორის, რადგან ბრაუნის კორპუსი არის პირველი თანამედროვე და ელექტრონული კორპუსი.

არანაკლებ მნიშვნელოვანია ბრიტანული ნაციონალური კორპუსი (British National Corpus - BNC), რომელიც დღესაც გამოიყენება და მიჩნეულია ინგლისური ენის რეპრეზენტაციულ კორპუსად. ბრიტანული ნაციონალური კორპუსის კონსორციუმი შეიქმნა 1990 წელს, რომელსაც ოქსფორდის უნივერსიტეტის გამომცემლობა ხელმძღვანელობდა. აღნიშნულ პროექტზე მუშაობა 1991 წელს დაიწყო და სამ წელიწადში 100 მილიონი თანამედროვე ბრიტანული ინგლისური სიტყვა შევიდა ბრიტანულ ნაციონალურ კორპუსში, აქედან 90% არის წერილობითი და 10% სასაუბრო. პირველი გამოცემა გამოვიდა 1994 წელს, ხოლო ევროპელი მკვლევრებისათვის ბრიტანეთის ეროვნული კორპუსი ხელმისაწვდომი გახდა 1995 წლის თებერვალში, კორპუსის რედაქტირებული ვერსია “BNC World“ გამოვიდა 2001 წელს, ხოლო 2007 წელს კი გამოვიდა მესამე გამოცემა. ბრიტანეთის ნაციონალური კორპუსი შედგება ორი ქვეჯგუფისაგან: 1) BNC Sampler (1 მილიონი წერილობითი და 1 მილიონი სასაუბრო ინგლისური) და 2) BNC Baby (4 მილიონი სიტყვა ოთხი განსხვავებული უანრიდან). ბრიტანეთის ნაციონალური კორპუსის მესამე გამოცემა მოიცავს 4049 ტექსტს, რაც შეადგენს

5,228,040 კილობაიტს, ან დაახლოებით 5.2 გეგაბაიტს. საერთო ჯამში კი მოიცავს 100 მილიონ სიტყვას, უფრო ზუსტად, 96,986,707 სიტყვას.

### 2.3.1 ქართული ენის კორპუსები

როგორც ზემოთ აღინიშნა, XX საუკუნის ოთხმოციანი წლებიდან მოყოლებული განვითარებული მსოფლიოს თითქმის ყველა ენისთვის, შეიქმნა ნაციონალური ენის კორპუსები, ვებკორპუსები, შეიქმნა რიგი კორპუსული პროგრამები და ინსტრუმენტები, რომლებიც ძალიან აქტიურად გამოიყენება დღეს კორპუსულ კვლევებში. რაც შეეხება ქართულ ენას, დღესდღეობით შექმნილია შემდეგი კორპუსები:

- 1) **TITUS - ARMAZI - კავკასიური ენები და კულტურები:** ქართულენოვანი ტექსტების პირველი აკადემიური ელექტრონული მონაცემთა ბაზა, რომელიც აერთიანებს სხვადასხვა პერიოდის ქართული ლიტერატურის ტექსტურ მასალას (<http://armazi.uni-frankfurt.de/framee.htm>). პროექტის ხელმძღვანელია ცნობილი გერმანელი ქართველოლოგი, ფრანკფურტის უნივერსიტეტის შედარებითი ენათმეცნიერების ინსტიტუტის დირექტორი, პროფესორი იოსტ გიპერტი, რომლის სახელსაც უკავშირდება ქართველური და კავკასიური ენების, განსაკუთრებით ძველი ქართული ენის ბეჭდური თუ ხელნაწერი მემკვიდრეობის უმდიდრესი ბაზების შექმნა;
- 2) **თანამედროვე ქართული ენის კორპუსი** - მოიცავს ორ სუბკორპუსს: თანამედროვე ქართული ენის კორპუსს (124,055,170 ერთეული) და ქართულ ლიტერატურულ კორპუსს, მორფოლოგიური ანოტირებით (20,903,850 ერთეული). პროექტის ხელმძღვანელია პაულ მოირერი ბერგენის უნივერსიტეტის უფროსი მეცნიერი თანამშრომელი.  
<http://clarino.uib.no/gekko/corpus-list>.

### 2.3.2 ქართული დიალექტური კორპუსი

საქართველოს ლინგვისტური პროგრეტი - ქართული დიალექტური კორპუსის პროექტს ხელმძღვანელობს არნ. ჩიქობავას ენათმეცნიერების ინსტიტუტის ენობრივ მონაცემთა კომპიუტერული დამუშავების განყოფილების ხელმძღვანელი, მთავარი მეცნიერი თანამშრომელი მარინე ბერიძე. ქართული დიალექტური კორპუსი წარმოადგენს ქართული ლინგვოკულტურული სივრცის ელექტრონული პროგრეტის შექმნის ცდას. საქართველოს ლინგვისტური პროგრეტი წარმოდგენილია სამი ქართველური ენითა (ქართული, სვანური, ზანური) და მათი დიალექტებით. კორპუსი თავისი სტრუქტურით იძლევა ქართული ენობრივი სინამდვილის სრულყოფილად წარმოჩენის შესაძლებლობას. კორპუსი შეიცავს არა მხოლოდ ტერიტორიული დიალექტების, არამედ ენის სხვადასხვა სინქრონიული თუ დიაქრონიული ვარიაციის აღწერის პოტენციურ შესაძლებლობასაც. კორპუსის შექმნის იდეა და დიალექტური მონაცემების შესწავლა თანამედროვე ტექნოლოგიების გათვალისწინებით ჩაისახა ჯერ კიდევ ადრეულ კვლევებში, რომლებიც კორპუსული ლინგვისტიკას ეფუძნებოდა და ეს საკმაოდ ბუნებრივი იყო. დიალექტოლოგია თავისი ბუნებიდან გამომდინარე ყოველთვის უფრო „კორპუსული“ იყო ენათმეცნიერების სხვა დარგებთან შედარებით, რადგან იგი ეფუძნება გარკვეულ შეგროვებულ ტექსტურ მასალას. თანამედროვე ქართული ლიტერატურული ენა ეფუძნება ქართველურ ენებსა და მის დიალექტებს;

ქართული დიალექტური კორპუსი – როგორც საძიებო-საინფორმაციო სისტემა შეიცავს მისი აქტიური გამოყენების შესაძლებლობებს მულტიმედიატექნიკური ხასიათის კვლევებისათვის. ელექტრონულად წარმოდგენილია კორპუსის საცდელი სამომხმარებლო ვერსია. კორპუსი ეყრდნობა პროექტის სამუშაო ჯგუფის მიერ მომზადებულ ელექტრონულ ტექსტურ ბაზას. ქართულ დიალექტურ კორპუსში ერთიანდება გასული საუკუნის ქართული დიალექტოლოგიის მონაპოვარი მთელი სიტყვიერი მასალა: ყველა, ხელმისაწვდომი და რამდენადმე ღირებული, დიალექტური ტექსტი; სალექსიკონო ერთეულები და საილუსტრაციო მასალები; დიალექტოლოგიურ ლიტერატურაში დამოწმებული ნიმუშები, შინამრეწველობის მასალები და ა. შ. თითოეულ ტექსტს ეძლევა დეტალური მეტადახასიათება. მრავალმხრივი მეტაანოტირების სისტემა შემუშავდა იმის გათვალისწინებით, რომ კორპუსის ინფორმაციული მონაცემთა ბაზა ადვილად გამოსაყენებელი ყოფილიყო

მულტიდისციპლინური კვლევებისათვის. კორპუსში ტექსტის აღსაწერად გამოყენებულია სხვადასხვა ექსტრალინგვისტური მახასიათებლები: რეგიონალური (მითითება რეგიონის, რაიონის, სოფლის, ენის, დიალექტის, ქვედიალექტის შესახებ), ანკეტური (ცნობები ტექსტის ჩამწერის შესახებ, მთქმელის სქესისა და ასაკის, დაბადების ადგილის, მასალის ჩაწერის დროისა და ადგილის შესახებ, მიგრაციული პროცესების ამსახველი მონაცემები), ტიპოლოგიური (ტექსტის ტიპი, თემატიკა), ქრონოლოგიური – ტექსტებში გადმოცემულ მოვლენათა პერიოდიზაციისათვის. მითითებულია ტექსტის სახე (ნაბეჭდი, ხელნაწერი, აუდიოტრანსკრიფტი, ვიდეოტრანსკრიფტი). ძიება ქართულ დიალექტურ კორპუსში მარტივი და მოსახერხებელია. მიუხედავად იმისა, რომ გრამატიკული ანოტირება დასრულებული არ არის, კორპუსის საძიებო სისტემა საჭირო ფორმების მოძიებისა და ლინგვისტური კვლევის საშუალებას ამ ეტაპზეც იძლევა. სეგმენტური ძიება (აფიქსის, ზმნისწინის, ნაწილაკის, ასობერის მიხედვით) უზრუნველყოფს ლექსემებისა და სიტყვაფორმების სრული კონკორდანსის მოწოდებას, კონტექსტებზე და ტექსტზე გასვლის შესაძლებლობებით (ბერიძე 2013).

### **2.3.3. ქართული ენის კორპუსი სიტყვათშეხამებითი**

#### **მოდელების პროგრამაში**

სიტყვათშეხამებითი მოდელების პროგრამა (Sketch Engine) შეიქმნა 2003 წელს შპს „ლექსიკალ კომპიუტინგის“ (Lexical Computing LTD) დამაარსებლისა და დირექტორის, კორპუსული და კომპიუტერული ლინგვისტის ადამ კილგარიფის მიერ. სიტყვათშეხამებითი მოდელების საძიებო სისტემა ვრცელი კორპუსული მონაცემების სწრაფად და ზუსტად დამუშავების შესაძლებლობას იძლევა სხვადასხვა კორპუსული ინსტრუმენტების გამოყენებით, რაც ძალიან მნიშვნელოვანია ლექსიკოგრაფიული საქმიანობისათვის. სიტყვათშეხამებითი მოდელების პროგრამას იყენებენ ლექსიკონების წამყვანი გამომცემლები როგორც დიდ ბრიტანეთში, ისე მის საზღვრებს გარეთ. დღეისათვის სიტყვათშეხამებითი მოდელების საძიებო სისტემა მოიცავს 200-ზე მეტ ენის კორპუსს, მათ შორის, ქართული ენის კორპუსს. სიტყვათშეხამებითი მოდელების

პროგრამაში ქართული კორპუსის შექმნა დაკავშირებულია დევიდ თავუელის სახელთან (Taguwell 2011). დღეისათვის, ქართულ კორპუსში დაახლოებით 24 მილიონი სიტყვაა, რომელიც ძირითადად ეფუძნება წერილობით ტექსტებს. უნდა აღინიშნოს, რომ კორპუსი ნაწილობრივ ანოტირებულია. სიტყვათშეხამებითი მოდელების საძიებო სისტემა არის კორპუსის ანალიზატორი, რომელიც არა მხოლოდ ავტომატურად აგებს სიტყვათშეხამებით მოდელებს, არამედ მასშტაბური კორპუსის ანალიზის მრავალფეროვან საშუალებებსაც მოიცავს:

- 1) **სიტყვათა ჩამონათვალი** (Word list), ძირითადი საძიებო (KWIC) ფუნქციები, კორპუსისა და ქვეკორპუსის შედარებები;
- 2) **განსხვავებები** (Sketch differences) - ორი სიტყვის განსხვავებულ მახასიათებლებზე დაკვირვება. ძიების ამ ინსტრუმენტის საშუალებით შესაძლებელია წარმოჩინდეს ის შემთხვევები, როდესაც სიტყვებს აქვთ საზიარო კოლოკაციები, ასევე შემთხვევები, როდესაც კოლოკაცია სიტყვათა წყვილიდან მხოლოდ ერთ-ერთისთვის არის რელევანტური;
- 3) **თეზაურუსი** (thesaurus) - თითოეული სიტყვისათვის ცალკეულად გამოითვლება დისტრიბუციული თეზაურუსი. ეს გვიჩვენებს კორპუსში სიტყვათა ჩამონათვალს, რომელიც გაწყობილია კონკრეტული სიტყვის რეალიზების ყველაზე ხშირი კონტექსტების მიხედვით;
- 4) **„ვებ-ბუთ-კეით“** (WebBootCat) – ეს არის სიტყვათშეხამებითი მოდელების პროგრამის ერთ-ერთი ძლიერი ინსტრუმენტი, რომლის საშუალებით შესაძლებელია ამა თუ იმ დარგობრივი/სპეციალიზებული კორპუსის შექმნა ძირეული სიტყვების ბაზაზე ავტომატურად (Baroni, Kilgarrieff, Pomikalek, Rychly 2006).

სიტყვათშეხამებითი მოდელების პროგრამაში განთავსებული ქართული ენის კორპუსისათვის ზემოთ ჩამოთვლილი ძიების შესაძლებლობიდან ამ ეტაპზე შესაძლებელია კონკორდანსების / სიტყვა-სიმფონიების (Concordance) მიხედვით ძიება, სიტყვანის (wordlist) ამოღება როგორც ანბანური რიგით, ასევე სისწორული განაწილების მიხედვით; შესაძლებელია ასევე „სიტყვის პორტრეტებისა“ (word sketch) და განსხვავებების (sketch diff) მიხედვით ძიება. გარდა ამისა, შესაძლებელია ამა თუ იმ სიტყვის თეზაურუსის (thesaurus) ნახვა.



### 2.3.4 ქართული ენის პარალელური კორპუსი

2013 წლიდან ივანე ჯავახიშვილის თბილისის სახელმწიფო უნივერსიტეტის ლექსიკოგრაფიულ ცენტრში მიმდინარეობს მუშაობა პარალელურ კორპუსზე (ინგლისურ-ქართული), რომლის შექმნა არიანე ჭანტურიას სამეცნიერო სტატიების რეზიუმეების პარალელური ქართულ-ინგლისური ტექსტებით დაიწყო.

ქართული ენის პარალელურ კორპუსში შეიქმნა ჯგუფები, რომლებიც დაიყო კრებულებად, ხოლო კრებულები ტექსტურ წყვილებად. კრებულები მოიცავს სამეცნიერო ჟურნალში გამოქვეყნებული სტატიების ინგლისურ-ქართულ თეზისებს. კორპუსის მეორე ჯგუფია არქეოლოგია, რომლის კრებულები მოიცავს: დმანისის გათხრების შესახებ გამოქვეყნებულ სტატიებს ინგლისური თარგმანებითურთ, ვანის გათხრებს და სხვა. კრებულებში შეტანილი ტექსტები ავტომატურად იშლება ტექსტურ წყვილებად, ინგლისურ-ქართულ წინადადებებად, რომლებიც რედაქტირდება ხელით, რათა ავტომატურად დანაწევრებული წინადადებები გათანაბრდეს და ძიების დროს მივიღოთ იდეალური ინგლისურ-ქართული წინადადებების წყვილები.

კორპუსში შეტანილი ტექსტები მეცნიერების თითქმის ყველა დარგს მოიცავს და ამ ტექსტებისა და დარგების რაოდენობა ყოველდღიურად იზრდება. დღეის მდგომარეობით კორპუსში არის 1908 კრებული, 28,000 ხელით გათანაბრებული ტექსტური წყვილი და 1 მილიონამდე სიტყვა-ფორმა.

კორპუსში დამუშავდა დარგების მიხედვით ტერმინოლოგიის ტაგირების / მონიშვნისა და კორპუსიდან აღნიშნული ტერმინოლოგიის ამოღების მეთოდოლოგია. ამ მიზნით შეიქმნა, ე.წ. ტერმინოლოგიური მოდული, რომელიც საშუალებას იძლევა კორპუსში არსებული მასალიდან ამოიკრიბოს წინასწარ მონიშნული დარგობრივი ტერმინოლოგია. ამ ეტაპზე ქართული ენის პარალელურ კორპუსის მართვის პანელში შესაძლებელია:

- კორპუსების ჯგუფების შექმნა და რედაქტირება;
- კრებულების შექმნა და რედაქტირება;
- ტექსტური წყვილების დამატება და რედაქტირება;
- ტექსტის ავტომატურ დანაწევრება, გათანაბრება და ტექსტური წყვილების გენერირება შემდგომი რედაქტირების შესაძლებლობით;
- ტექსტში ტაგირებული / მონიშნული ტერმინების აღქმა და ძიება კორპუსში.

აქვე უნდა აღინიშნოს, რომ სამეცნიერო ტექსტების ინგლისურ-ქართულ პარალელურ კორპუსს უკვე აქვს საძიებო ინტერფეისი, მათ შორის, დეტალური ძიების ინტერფეისი დარგობრივი ტერმინოლოგიისათვის, რომელიც განკუთვნილია კორპუსის მონაცემთა ბაზაში სასურველი ქართული ან ინგლისური სიტყვის ან შესიტყვების ძიებისა და შესაბამისი შედეგების შემცველი პარალელური ტექსტური წყვილების გამოტანისთვის. ძიება ხორციელდება ორმხრივად - როგორც ნაწარმოების ორიგინალ, ასევე, ნათარგმნ ტექსტში. ძიება შესაძლებელია როგორც მთლიან მასალაში, ასევე, სხვადასხვა პარამეტრის მიხედვით, მათ შორის, ნაწარმოების ჯგუფების, კრებულების, ტიპების, ჟანრების, ავტორების, წლების და ა.შ. ფილტრაციით. ძიების შედეგებში ასევე, ხელმისაწვდომია ინფორმაცია ნაწარმოების შესახებ, მათ შორის, ავტორის, თარგმანის ავტორის, გამოცემის წლისა და ა.შ. შესახებ (თ. მარგალიტაძე, ქერეჭაშვილი 2013).

### 3. ვეზი როგორც კორპუსი

#### 3.1 ზოგადი მიმოხილვა

წინამდებარე ქვეთავში მიმოვიხილავთ კორპუსული ლინგვისტიკის ახალ მიმართულებას, რომელიც ეხება ვეზის ლინგვისტურ კორპუსად გამოყენებას.

არის თუ არა ვეზი კორპუსი?

ამ შეკითხვაზე რომ ვუპასუხოთ, უპირველეს ყოვლისა, უნდა გავიგოთ რა არის თვითონ კორპუსი? მაკენერისა და უილსონის მიხედვით კორპუსი შეიძლება ვუწოდოთ ერთზე მეტი ტექსტის კრებულს (McEnery & Wilson 1996), თუმცა ენათმეცნიერებაში „კორპუს“ თავისი აუცილებელი პარამეტრი და საზომი გააჩნია. თანამდროვე ენათმეცნიერული გაგებით კორპუსს გააჩნია ოთხი ძირითადი პარამეტრი: შერჩევა და რეპრეზენტაციული, სასრული, ელექტრონულად წაკითხვადი, სტანდარტული. აქვე შეიძლება ითქვას, რომ ძალიან ბევრი ტექსტების კრებული, რომლებსაც კორპუსი ეწოდება, ვერ აკმაყოფილებს ზემოთ ჩამოთვლილ პარამეტრებს. ადამ კილგარიფის მიხედვით, კორპუსი, რომელიც მოიცავს ჯეინ ოსტინის მთლიან შემოქმედებას, არ შეიძლება იყოს რეპრეზენტაციული რაიმეს მიმართ (Kilgarriff 2003). მანინგისა და შუცეს მიხედვით კი სტატისტიკური, ბუნებრივი ენის დამუშავებაში (Statistical NLP) კორპუსად განიხილება ტექსტური მონაცემების გარკვეული რაოდენობა კვლევის გარკვეული ინტერესისათვის, მიუხედავად იმისა, თუ როგორ და რა პრინციპით არის კორპუსი შედგენილი. ასეთ შემთხვევაში, დამუშავებული და გაფილტრული კორპუსული მასალა ბევრად უფრო მნიშვნელოვანია, ვიდრე კორპუსის დაბალანსება (Manning & Schütze 1999).

მაკენერი და უილსონი (McEnery & Wilson 1996) ერთად განიხილავენ ორივე შეკითხვას: „რა არის კორპუსი“ (“what is a corpus?”) და „რა არის კარგი კორპუსი“ (“what is a good corpus?”) და აღნიშნულ საკითხს შედარებით მარტივი მაგალითით ხსნიან: „არის თუ არა  $X$  კორპუსი კარგი  $Y$  ამოცანის გადასატრედად?“ (“is corpus  $X$  good for task  $Y$ ?”), ამის საპირისპიროდ მათ შემოაქვთ, ე.წ. „სემანტიკური“ შეკითხვა, „არის კი  $X$  კორპუსი?“ (“is  $X$  a corpus at all?”), როგორც ვხედავთ, „სემანტიკურ“ შეკითხვას ყურადღება სხვა რამეზე გადააქვს, უფრო ნაკლებად მნიშვნელოვან საკითხზე, მაშინ, როდესაც კორპუსი უნდა განვიხილოთ პირველი შეკითხვის ჭრილში – „არის თუ არა  $X$  კორპუსი კარგი  $Y$  ამოცანის

გადასაჭრელად?“ (“is corpus X good for task Y?”). გამომდინარე ზემოაღნიშნულიდან, მაკენერი და უილსონი გვთავაზობენ „კორპუსის“ დეფინიციას, რომელიც ფორმულირდება შემდეგნაირად: **„კორპუსი არის ტექსტების კრებული, როდესაც იგი განიხილება ენისა და ლიტერატურის კვლევის ობიექტად“** (“a corpus is a collection of texts when considered as an object of language or literary study”), ამგვარად, პასუხი კითხვაზე, არის თუ არა ვები კორპუსი, დადებითია (Kilgarriff & Grefenstette 2003).

### 3.2 ვებკორპუსები

დღითი დღე იზრდება ინტერესი ვების მიმართ და იგი წარმოადგენს ყველაზე ვრცელ ლინგვისტურ წყაროს. ვების ბაზაზე შექმნილ კორპუსებს ეწოდება ვებკორპუსები, რომლებიც კორპუსების ახალ და დიდ ჯგუფს წარმოადგენენ.

ვებკორპუსის შექმნის პირველ მცდელობას წარმოადგენს 53-მილიარდსიტყვიანი ტერაბიტების კორპუსი, რომელიც ვატერლოოს უნივერსიტეტში შეიქმნა (Clarke et al. 2002). აღნიშნული კორპუსი ეფუძნება უნივერსიტეტების და სხვა საგანმანათლებლო დაწესებულებებისა და ორგანიზაციებიდან ჩამოტვირთულ ინტერნეტმასალას. ტერაბიტების კორპუსის ანალოგს წარმოადგენს “GOV2 კორპუსი”, რომელიც ეფუძნება “.gov” დომენს (Clarke et al. 2005).

ვებკორპუსების ერთ-ერთ საინტერესო მაგალითად შეიძლება განვიხილოთ თელვალის ვებკორპუსები, რომლებიც ეფუძნება ავსტრალიაში, ახალ ზელანდიასა და დიდ ბრიტანეთში აკადემიური ხასიათის ვებგვერდებიდან ჩამოტვირთულ ინტერნეტმასალას (Thelwall 2005). აქედან, ყველაზე ვრცელია დიდი ბრიტანეთის ვებსაიტებიდან შეგროვებული კორპუსი, რომელიც დაახლოებით 1.5 მილიარდ სიტყვას მოიცავს. თელვალს დეტალურად აქვს აღწერილი ამ კორპუსის ჩამოტვირთვასთან, ვებგვერდების ნახევრად ავტომატურ დამუშავებასთან და ა.შ. დაკავშირებული საკითხები (Thelwall 2005).

სისწორული ლექსიკის განსაზღვრისათვის ვებკორპუსები გამოიყენა კორნაიმ (Kornai et al. 2006), რომელიც მუშაობს უნგრულ ენაზე, ასევე ემერსონმა და ონილმა (Emerson & O'Neil 2006), რომლებიც ჩინური ენის ვებკორპუსზე მუშაობენ. განსაკუთრებით შთაბეჭდავია „გუგლის ტერაბაიტების ენგრამების კორპუსი“ (Google Terabyte n-gram Collection), რომელიც შეიქმნა 2006 წელს (Brants & Franz 2006).

ლიუმ და კურანმა შექმნეს 10-მილიონსიტყვიანი ინგლისური ენის ვებკორპუსი (Liu & Curran 2006), რომელიც სხვადასხვა ვებსაიტებიდანაა შერჩეული ღია საძიებო სისტემისათვის (Open Directory Collection). ლიუსა და კურანის ვებკორპუსი წინადადების დონეზეა ჩაშლილი, გაფილტრული და ტოკენიზებული.

არანაკლებ საინტერესოა შაულისა და ვესტბურის ვებკორპუსი „იუნეტი“ - USENET (Shaoul & Westburry 2007), აღნიშნული კორპუსი 2005-2007 წლებში იქმნებოდა და მოიცავს 13 მილიარდ სიტყვას. გარდა ამისა, უნდა აღინიშნოს, კატალანიურის ვებკორპუსი – CUCWeb (Boleda et al. 2006), რომელიც მოიცავს 166 მილიონ სიტყვას, კატალანიური ენის კორპუსი ვებკორპუსის შექმნის ყველა პარამეტრების გათვალისწინებით განხორციელდა, როგორცაა გაფილტვრა, დუბლიკატების მოხსნა, მორფოსინტაქსური ანოტირება, გარდა ამისა, აღნიშნული ვებკორპუსი ხელმისაწვდომია და გააჩნია ადვილად გამოსაყენებელი ინტერფეისი მომხმარებლებისათვის (<http://www.catedratelefonica.upf.es/cucweb>).

სერგეი შაროვმა შეიმუშავა ბრიტანული ნაციონალური კორპუსის ზომის (“BNC-sized) ვებკორპუსები, რომლებიც დაახლოებით 100 მილიონ სიტყვას ეფუძნება (Sharoff, 2006). სერგეი შაროვის მიერ შექმნილ ვებკორპუსებს შორის უნდა აღინიშნოს ინგლისურის, ჩინურის, ფინურის, ფრანგულის, გერმანულის, იაპონურის, პოლონურისა, პორტუგალიურის, რუსულისა და ესპანურის ვებკორპუსები, ამ ჩამონათვალს დაემატა ქართული ენის ვებკორპუსი, რომელიც წინამდებარე ნაშრომის თემას წარმოადგენს და, როგორც ზემოთ აღვნიშნეთ, სადოქტორო კვლევის ნაწილი ლიდის უნივერსიტეტში განხორციელდა სერგეი შაროვის მეთოდოლოგიის გამოყენებითა და უშუალოდ მისი მონაწილეობით.

გარდა ზემოთ ჩამოთვლილი ვებკორპუსებისა, უნდა აღინიშნოს WaCky-ის კორპუსები (Baroni, Bernardini, Ferraresi and Zanchetta 2009). WaCky წარმოადგენს Web as Corpus kool ynitiativ-ის პროექტის ერთ-ერთ ნაწილს. აღნიშნულ პროექტში

ჩართული იყო მეცნიერთა კონსორციუმი, რომელთა ინტერესის სფეროს სწორედ ვები და მის ლინგვისტურ წყაროდ გამოყენებასთან დაკავშირებული საკითხები წარმოადგენდა. WaCky-ის პროექტის ფარგლებში შექმნილია შემდეგი ენის ვებკორპუსები:

- 1) ინგლისური ენის ვებკორპუსი – **UkWaC**, კორპუსი მოიცავს 2 მილიარდ სიტყვას, ძირითადად **.uk** დომენიდან და იყენებს ბრიტანული ნაციონალური კორპუსის მიხედვით საშუალო სიხშირის სიტყვებს საკვანძო სიტყვებად (seed words), კორპუსი მორფოლოგიურად ანოტირებულია და ლემატიზებულია. **UkWaC** კორპუსი ძიებისათვის ხელმისაწვდომია სიტყვაშეხამებითი მოდელების პროგრამაში: <https://www.sketchengine.co.uk/documentation/wiki/Corpora/UKWaC>;
- 2) ინგლისური ენის ვებკორპუსი **PukWaC**, იგივეა, რაც **UkWaC**, მაგრამ განსხვავება მდგომარეობს ანოტირების ნიუანსში, კერძოდ, კორპუსი სრულად არის პარსირებული. პარსირება განხორციელდა სპეციალური პარსირების ინსტრუმენტის საშუალებით, რომელსაც MaltParser (<http://maltparser.org/>) ეწოდება (Johansson 2007);
- 3) ინგლისური ენის ვებკორპუსი – **WaCkypedia\_EN**, წარმოადგენს 2009 წლის ინგლისური ვიკიპედიის კრებულს (დაახლოებით 800 მილიონ ტოკენს მოიცავს), აღნიშნული კორპუსი წარმოდგენილია **PukWaC**-ის ფორმატში. იგი ლემატიზებული და პარსირებულია. ტექსტების ჩამოტვირთვა და გაწმენდა განხორციელებულია სპეციალური ინსტრუმენტის Wikipedia extractor-ის საშუალებით;
- 4) ფრანგული ენის ვებკორპუსი – **frWaC**, მოიცავს 1.6 მილიარდ სიტყვას, ტექსტები ძირითადად ჩამოტვირთულია **.fr** დომენიდან და იყენებს საშუალო სიხშირულ სიტყვებს საკვანძო სიტყვებად. საშუალო სიხშირის საკვანძო სიტყვები აღებულია ფრანგული ენის დიპლომატიური კორპუსიდან (Le Monde Diplomatique Corpus) და ფრანგული ენის ძირითადი ლექსიკიდან (basic French vocabulary lists). კორპუსი მორფოლოგიურად ანოტირებულია და ლემატიზებულია (Baroni, Bernardini, Ferraresi, Zanchetta 2009);
- 5) გერმანული ენის ვებკორპუსი – **deWaC**, მოიცავს 1.7 მილიარდ სიტყვას. ტექსტები ძირითადად ჩამოტვირთულია **.de** დომენიდან და იყენებს

საშუალო სიხშირულ სიტყვებს საკვანძო სიტყვებად. საშუალო სიხშირის საკვანძო სიტყვები აღებულია გერმანული ენის კორპუსიდან (SudDeutsche Zeitung corpus) და გერმანული ენის ძირითადი ლექსიკიდან (basic German vocabulary lists);

6) გერმანული ენის ვებკორპუსი – **SdeWac**, კორპუსი მოიცავს 0,88 მილიარდ სიტყვას **deWaC**-ის ბაზაზე. **SdeWac** გადაყვანილია სპეციალურ უნიკოდში და ხელმისაწვდომია კორპუსის ორი ვერსია, პირველი ვერსია მორფოლოგიურად ანოტირებული და ლემატიზებულია, ხოლო მეორე ვერსია კი წინადადების დონეზეა ანოტირებული შესაბამისი მეტაინფორმაციის მითითებით, მაგალითად, პარსირების შეცდომის მითითებით.

7) იტალიური ენის ვებკორპუსი – **itWaC**, მოიცავს 2 მილიარდ სიტყვას. ტექსტები ძირითადად ჩამოტვირთულია **.it** დომენიდან და იყენებს საშუალო სიხშირულ სიტყვებს საკვანძო სიტყვებად. საშუალო სიხშირის საკვანძო სიტყვები ეფუძნება იტალიური ენის კორპუსსა (Repubblica) და იტალიური ენის ძირითად ლექსიკას (basic Italian vocabulary lists). იტალიური ენის ვებკორპუსი მორფოლოგიურად ანოტირებულია, ხოლო ლემატიზება განხორციელებულია სპეციალური მორფემების ლექსიკონის საშუალებით (Morph-it! lexicon).

ეჭვგარეშეა, ვებკორპუსების მნიშვნელობა თანამედროვე კორპუსული ლინგვისტიკაში, რაზეც ვებკორპუსების მზარდი რიცხვი მეტყველებს. დღითი დღე იზრდება ვებკორპუსების რაოდენობა, იქმნება რიგი კორპუსული ინსტრუმენტებისა და იზრდება ვებკორპუსების მოცულობა და იხვეწება მათი შექმნისა და გამოყენების მეთოდოლოგია. ამ ფონზე, აშკარაა ქართული ენის მდგომარეობა, ქართული ენის ვებკორპუსი – **KaWaC**, რომელიც აღნიშნული სადოქტორო კვლევის ფარგლებში შეიქმნა, არის პირველ მცდელობა ამგვარი კორპუსის შექმნისა ქართული ენისათვის ქართული ენის ვებკორპუსსა და მის შექმნასთან დაკავშირებულ საკითხებს, ქვემოთ, II თავში განვიხილავთ.

\*\*\*

ამგვარად, სადისერტაციო ნაშრომის პირველ თავში მიმოვიხილეთ სასწავლო ლექსიკოგრაფიის განვითარების ისტორია, პირველი სასწავლო ლექსიკონები, ინგლისური ენის სასწავლო ლექსიკონების „დიდი ხუთეული“: ოქსფორდის

სასწავლო ლექსიკონი, ლონგმენის თანამედროვე ინგლისური ენის ლექსიკონი, კოლინზ ქობილდის ინგლისური ენის ლექსიკონი, კემბრიჯის ინგლისური ენის საერთაშორისო ლექსიკონი; მაკმილანის ინგლისური ენის ლექსიკონი. ამავე თავში წარმოდგენილია და განხილულია კორპუსული ლინგვისტიკა და კორპუსები. კერძოდ, კორპუსული ლინგვისტიკის განვითარების ეტაპები, კორპუსის რაობა, კორპუსის ტიპები და ენათა ისეთი კორპუსები, როგორცაა: ბრაუნის კორპუსი, ბრიტანული ნაციონალური კორპუსი, ასევე ქართული ენისათვის შექმნილი კორპუსები: TITUS - ARMAZI - კავკასიური ენები და კულტურები, თანამედროვე ქართული ენის კორპუსი, ქართული დიალექტური კორპუსი, ქართული ენის კორპუსი სიტყვათშეხამებითი მოდელების პროგრამაში, ქართული ენის პარალელური კორპუსი. გარდა ამისა, მიმოხილულია ახალი ტიპის კორპუსები – „ვები როგორც კორპუსი“, მისი განვითარების ისტორია და აღწერილია ისეთი ენის ვებკორპუსი, როგორცაა **UkWaC** - ინგლისური ენის ვებკორპუსი, **frWaC** - ფრანგული ენის ვებკორპუსი, **deWaC** გერმანული ენის ვებკორპუსი და ა.შ.



## II თავი. ქართული ენის ვებკორპუსი – KaWaC

### 1.1 ზოგადი მიმოხილვა

წინამდებარე თავში განვიხილავთ ქართული ენის ვებკორპუსს, მის შექმნასთან დაკავშირებულ ისეთ საკითხებს, როგორცაა კორპუსის ტექსტუალური მასალის შერჩევა, ჩამოტვირთვა, მორფოლოგიური ანოტირება და ა.შ. უნდა აღინიშნოს, რომ ქართული ენის ვებკორპუსი შეიქმნა სადოქტორო პროგრამის ფარგლებში ლიდსის უნივერსიტეტში ამავე უნივერსიტეტის პროფესორის სერგეი შაროვის უშუალო მონაწილეობით. კორპუსი განთავსებულია ლიდსის უნივერსიტეტის ვებსაიტზე და მოიცავს 150 მილიონ სიტყვას (<http://corpus.leeds.ac.uk/internet.html>).

### 1.2 ქართული ენის ვებკორპუსი - KaWaC

ქართული ენის ვებკორპუსი – **KaWaC** იშიფრება შემდგენიერად: **Ka** - **Kartvelian** - ქართული და **WaC** – **Web as a Corpus** – ვები, როგორც კორპუსი, ე.ი. ქართული ენის ვებკორპუსი. აღნიშნული კორპუსი შეიქმნა ლიდსის უნივერსიტეტში სამთვლიანი სადოქტორო პროგრამის კვლევის ფარგლებში.

ქართული ენის ვებკორპუსი ასახავს ბოლო 10-15 წლის ქართულ ენას, ამ პერიოდის განსხვავებულ უანრულ თუ რეგიონალურ ვარიაციებს (რეგიონალურ ვარიაციში იგულისხმება არა დიალექტური მეტყველება, არამედ ვებრესურსები საქართველოს სხვადასხვა რეგიონიდან, მაგალითად, როგორცაა თელავის, ზუგდიდის, გორის სასწავლო უნივერსიტეტები, ადგილობრივი ტელევიზიები/პრესა და ა.შ.) და მოიცავს 150 მილიონ სიტყვას, უფრო ზუსტად:

- ტექსტი: 618468 ვებგვერდი - 697 ვებსაიტიდან
- სიტყვა: 182845341, 2865042 ტიპი
- ტოკენი: 23550807, 3447266 ტიპი

ზემოთ „სიტყვაში“ იგულისხმება ცალკეული სიტყვაფორმის ხმარების შემთხვევა, ხოლო „ტიპში“ განსხვავებული/უნიკალური სიტყვაფორმები. ტოკენი კი მოიცავს როგორც სიტყვაფორმებს, ასევე პუნქტუაციის ნიშნებს.

როგორც უკვე ითქვა, KaWaC განთავსებულია ლიდსის უნივერსიტეტის ვებგვერდზე (იხილეთ სურათი 1) და აღჭურვილია ონლაინ საძიებო სისტემით, ამ ეტაპზე შესაძლებელია სიტყვანის (Word List), კონკორდანსების ძიება

კორპუსში და შესაბამისი კონტექსტის ნახვა. კორპუსში შესაძლებელია ძიება მეტყველების ნაწილების მარკერების – POS Tag (Part-of-Speech Tag) მიხედვით. გარდა ამისა, ჩვეულებრივი (არა მეტყველების ნაწილების მარკერების) ძიების შემთხვევაშიც გამოტანილია ამა თუ იმ სიტყვის შესაბამისი მეტყველების ნაწილის მარკერი და ლემა (იხილეთ სურათი 2).

### სურათი 1: ქართული ენის ვებკორპუსი – KaWaC

The screenshot shows the KaWaC web interface. At the top, there is a search bar with the text "Search" and several utility icons: "from DOC to PDF", "Convert to PDF", "Convert to DOC", "Translate", "Thesaurus", "Dictionary", and "Reference". Below the search bar, there is a link "Back to the query window". The main content area displays the search results for the word "გული". The title is "See 70 examples of '[word='გული'] cut 100' in KAWAC". The results are listed as a series of text snippets, each preceded by a blue double arrow icon (»). The snippets show various contexts where the word "გული" is used, such as "სტრუქტურას და ელასტიურობას. მას შეიცავს სტაფილო, კვერცხის გული, თევზული. ვიტამინი E დაგეხმარება თმების სწრაფ გაზრდაში, ... ვალში, ვერ ამოვალ მასთან ... სახლში, მყუდროდ არის სახლში ... გული, თავალს კი მაინც რული ... ადარ სტუმრობს ... რული ... სვედით სავესე არ ჰქონდეს ავტორიტეტი. არ გამოვა ასე, მე ძალიან დამწყდება გული, თუ იმ საქმეებს, რაც ბიძიან ივანიშვილს ქვეყნისთვის ორგანიზმის შესაცვებად კი კარგია რზის სხივები, კვერცხის გული, ორგული, რომელიც ასევე ომეგა-3-ის ცხიმოვანი მჟავებითაა და შეიცავს ანტიოქსიდანტებს ( განსაკუთრებით, კვერცხის გული, რომელშიც შედის ლუტეინი ). ის გემრიელი საუზუნა და . სადღაც ბღავის ფური ... თიხას გახეთქს, გამოჟონავს ... და მაინც გული, ცდილობს იყოს მასთან ... ვინემ, ამოიღოს სული. ვერ ელევა, სურს იყო, როგორც ბოლო წლებში. ბარსელონას ყველა შეტევის სული და გული . <პ> ანდრესზეც ახალი რა უნდა ვთქვათ, გარდა იმისა, რომ მსგავსი რაღაც იყო და ისიც მირითადად უშედეგო. ძალიან დამწყდა გული . <პ> ახლა კი 2008 წელია. ოთხი წელი გავიდა იმ ამბების მხოლოდ ერთმა მოიგო, სამაგიეროდ ორივემ მოიგო გულშემატკივრის გული . <პ> ფერგიუსონს არ ბეზრდება გამარჯვებები. მეცხრამეტე ასეთი სიტუაითი და სიცვარულით არის განათებული ამ ადამიანების გული . ბოლო დღეებში ძალიან ბევრი შეხვედრა მქონდა, ყველგან მიჩდიდა იმ ეპარქიისა, სადაც მტკივნეულად ფეთქავს მთელი საქართველოს გული . ვულოცავ მის უწინდესობას ამ თარიღს და ვუსურვებ დიდხანს იმ ეპარქიისა, სადაც მტკივნეულად ფეთქავს მთელი საქართველოს გული . თქვენო უწინდესობავ, გილოცავთ დაბადების 78-ე წლისთავს. " შედეგად მას აუცილებლად დაუზიანდება სისხლძარღვები და გული ; ბ ) თავდაცვის სამსუალების ( რქებისა და ეშვების ) უქონლობის ბრინჯის ფაფით სავსე ჯამში. სახალწლო კერძებში კი თხილის გული ა მოთავსებული. ქალიშვილი, რომელიც ამ თხილის გულს იპოვის, ცდილობენ დამახინჯებული ადამიანების დემონსტრირებით მასას გული არევიწოს და გაუმბაფროს ჰომოფობია ქსენოფობია. <პ> - თუ ბეღელ დაგაღებებს თავს ისე მარტივად ვეღარ ართმევთ და გული ადვილად იტვირთება, ეს უკვე წითელი დროშის ნიშანია, რაც იმას ერთსაათიანი დატვირთვა სასიცოცხლოდ აუცილებელია. წავარჯიშები გული ადვილად უშკლავდება გამოიერებული რიტმით მუშაობას. ფიზიკური არ ჩატუტარების სხეულისგან განსხვავებით ( იმ ე. წ. ორსულზე გული ამერია: @ ), ბელგრადელი გოგონა მორიდებულია და საერთოდ ? p=1163#comment-236 არ ვმალავ რომ სენტემენტი ვარ, ბილოსკენ გული ამიჩუყდა. სტუდენტობის დაწყებისას ვხვდებით რომ ბავშვობა ერთი იყო ვინმე ამაკო სალუქვადე, რომლის სიკვდილზე გლეხებს გული არ დასწყევტიათ, რადგან თურმე ის მოკლულ კაზაკს ჩემებს ხდიდა "ვედივარ" ... <პ> მიდის ... <პ> წავიდეს, მაგრამ გული არ მატკინოს ... არაუშავს ... <პ> და ისევ სავარძელთან სანამ ის ბიძის მომლო ხინშილი გააოჯებდა, ლამის იყო გული არეოდა. გათავებას ბიურმანმა ისე ძლიერად წაუჭირა ყელში

## სურათი 2: საძიებო სისტემა ქართული ენის ვებკორპუსში

[Back to the query window](#)

See 738 examples of '[word="ქალმა"] cut 1000' in KAWAC

<p>შენი თავგადასავალი მახსოვს. რა უნდა მამაკაცს, რომ იცოდეს უბიძგებს. • ზოგს არ სიამოვნებს, რომ ის ძალისხმევა, რომელიც , სამი " ვერცხლის წმ. გიორგი " ერთადერთმა ფილმმა - " ფრანგმა ? მას არასდროს უნახავს ჩემი ცრემლები. - ჩურჩულით თქვა განმავლობაში ეს ხარკები საკუთარ თავზე აიღო. ნ. ბრიტანელმა ქალაქზე ამბავი ( რომელსაც ვერ შეამოწმებ ) თუმცა რაც ამ მ მარტს ქალთა დღესთან დაკავშირებით გამოსულმა ქართველმა . ქართა ველიდან მოსულები. ქარვით. ცხელი ქარვით ... და სხვებისთვისაც სასარგებლო, რა უნდა გააკეთო კაცმა ( ან მე არ ვამბობ, რომ ადრეულ ასაკში უნდა დაიწყო ადამიანმა საუბარი არჩია. ნიმ ანუ სიყვარულმა ( თუ საყვარელ პირველი მსახიობი ტრანსსექსუალია. მოხდენილმა აზიელმა მოხუცმა, ნეტავ თუ იპოვა სიმშვიდე? ' ' იკითხა ერთმა ფამფამა და ჩემი დაქალი ავტობუსით ვმგზავრობდით. ჩვენ წინ მჯდომმა სიტყვებისთვის. - ალბათ, ცუდ საპონს იყენებს, - გააგრძელა მშრალი რომ ბრძანდებით. III " III " თქვენ კი, - განაცხადა , მაგალითად, პაატას. აი, შენ კი, - სიტყვა გააგრძელა უნდა. - უკაცრავად, ახლა არ მცალია, - მშვიდდით უპასუხა ... მალე დაბრუნდება? - ანა, რა გითხრათ, თავი გაიქცია ჰქონდათ. კაცმა დვინო შესთავაზა: - არა, მვირფასო, - თქვა ძალიან ლამაზი ხართ _ თქვა მან _ ძალიან ლამაზი? _ გაიკვირვა არის ახლა, კაცმა სიყვარული მალოს, ადარც, სხვათაშორის, - დააგლო და ზურგი შეატყია. იმავე დღეს ელუნესთან გადაბარდა. კომპლექსად ჩამითვალეთ მაგრამ ვთვლი, რომ ორმა მეგობარმა ამ დროს ძლიერად დააკაუნეს კარზე. ჭურჭელს თავი მიანება . მაგანაც ვერც ქნა, მაგ საწყალმა. ერთი შვილი იქვე გაუჩინა წავაწყდით პირველ პრობლემას, ეს არის მარტოხელა დედის ტვირთი.</p>	<p>ქალმა   1. გამოხატე შენი აზრი. მე ამაყი ვარ ამ დროს, თუ გინდა   12:53 / 01-11-2012 ნიკა ნათია-ს, ნათი ცემო კარგო, არ გინდა ქალმა " ბიჭის " მამაკაცად ჩამოყალიბებაში ჩადო, სხვა ქალს ქალმა " მიიღო. ბედის ირონიით, იმ მომენტში, როცა ბუარი საფრანგეთის ქალმა " ოპ, ეს სუსტი სქესი: ) -მემინია, ხელახლა არ შემოვივარდე. ქალმა " ტვიტერში " იმზობიარა დიდი ბრიტანეთის 34 წლის მცხოვრებმა " (არქეოლოგმა) ილაპარაკა, თუ ყველაფერი მართალია, ამის ქალმა (და არამარტო) პოლიტიკოსებმა. კიდევ, " ქალთომოძულე კაცებმა ქალმა (თანაც ლამაზმა და თანაც ვარსკვლავმა) ასაკი რომ დამალოს, ქალმა ) თანაც, მიუხედავად იმისა, რომ უამრავი რამ ჩამოთვალე, ქალმა ქალმა/Ncse ) ბიჭი ცხოვრება, არა ეს ცუდია ჯანმრთელობისთვის. ქალმა ) თორღო იმდენად იგი გამოველინა თავისი სულიერება. სიყვარულის ქალმა , " მამაკაცური " წარსულის მიუხედავად, Miss International ქალმა , " გააჩნია სიმშვიდეს რას დაარქმევ? " აღმოხდა მაშინ მოსილ ქალმა , " უი, დედიკოს როგორ გავსო " და ლენას კალთაში მჯდარ სერიოფას ქალმა , - ან კიდევ საერთოდ არ იცი რეცხვა. ქართველებს, სხვა ხალხთა ქალმა , - გირჩევნიათ, გალუმპულუბი იდგეთ. III " შერე ფანჯარა ქალმა , - ნამდვილად იმსახურებ გენიოსს, თანაც ყოველთვის ოცნებობდით ქალმა , - როგორც კი ცოლის მოყვანას გადაწყვეტ, მაშინვე წავალ ქალმა , - არასდროს იმარებს, როდის მოვა. ვინ ბრძანდებით? თუ გინდათ, ქალმა , - მე მეტი აღარ მინდა. კაცმა კვლავ სთხოვა: - მხოლოდ თითო-თითო ქალმა , აკისკიცდა და ხელი ხელზე დაადო. ალბათ ხედები, მკითხველო, ქალმა , ამ მხრივ დღეს არავინ წვალობს!. არცა, მეგობრების გარდა, ქალმა , არ იცოდა, სიხარულით რა ვკეთებინა. სადღაც, გულის სიღრმეში, ქალმა , არ უნდა იცოდეს ერთი და იგივე მამაკაცის სექსუალური ქალმა , აუჩქარებელი ნაბიჯით მივიდა გასასვლელთან. - უკაცრავად, აქ ქალმა , ბიჭი, ჯარში, მეორეც, გოგო, მაშინვე მიუყვალა ფათიფითით. ქალმა , გარდა იმისა, რომ მატერილურად უნდა უზრუნველყოფს შვილები,</p>
--	---

მე-2 სურათზე ნაჩვენებია კორპუსის საძიებო სისტემა, ძიების ველში საძიებო სიტყვად გამოტანილია სიტყვა-ფორმა „ქალმა“, რასაც თან ახლავს მოცემული სიტყვის შესაბამისი მარკერი და ლემა. მარცხნივ მოცემულია მარკერი, ე.წ. “POS ტაგი - “Ncse” (ე.ი, *Noun, Type = Common, Number = Singular, Case = Ergative*) და მოიცავს შემდეგ ინფორმაციას: *არსებითი სახელი, ტიპი = საზოგადო, რიცხვი = მხოლოდობითი, ბრუნვა = მოთხრობითი*, ხოლო მარჯვნივ კი მოცემულია ამ სიტყვის ლემა „ქალი“.

### 1.3 ვებგვერდების ლინგვისტური დამუშავება

ქართული ენის ვებკოპუსის შექმნის პროცესში პირველ ეტაპს წარმოადგენდა კორპუსის ტიპის განსაზღვრა და კორპუსისათვის შესაბამისი მასალის შერჩევა. ქართული ენის ვებკოპუსში რეპრეზენტაციულობის მიღწევის მიზნით შეირჩა 697 ვებსაიტი, რაც ჯამში 618468 ვებგვერდს შეადგენს. ვებგვერდები პროგრამა wget-ის (World Wide Web და get) გამოყენებით ჩამოიტვირთა. Wget არის კომპიუტერული პროგრამა, რომელიც ვებსერვერებიდან ტექსტური მასალის ჩამოტვირთვის შესაძლებლობას იძლევა და წარმოადგენს GNU პროექტის ნაწილს, აღნიშნული პროექტი ითვალისწინებს კომპიუტერული ინსტრუმენტების ხელმისაწვდომობას (Stallman 2003). უნდა აღინიშნოს, რომ უშუალოდ ვებგვერდების ჩამოტვირთვამდე განისაზღვრა 300 ფუნქციური სიტყვა (function words) ქართული ენისათვის. ფუნქციური სიტყვები ამა თუ იმ ენის ყველაზე ხშირად ხმარებულ ლექსიკას წარმოადგენს და ქართული ენის ფუნქციური 300 სიტყვა შეირჩა დევიდ თაგუელის ქართული ენის კორპუსის მიხედვით, რომელიც სიტყვათშეხამებითი მოდელების პროგრამაშია განთავსებული. Wget-ს პროგრამაში განთავსდა ქართული ენისათვის შერჩეული 300 ფუნქციური სიტყვა (იხილეთ სურათი 3). ფუნქციური სიტყვების განსაზღვრა უშუალოდ ვებგვერდებიდან ტექსტური მასალის ჩამოტვირთვამდე ძალიან მნიშვნელოვან ეტაპს წარმოადგენდა კორპუსის შექმნის პროცესში, რათა გამოგვერიცხა არაქართული გვერდები, ანიმაციები და სხვა ლინგვისტურად არარელევანტური ინფორმაცია.

ამგვარად, 697 ვებსაიტიდან ჩამოიტვირთა სულ 618468 ვებგვერდი. კორპუსის შექმნის შემდეგ ეტაპს წარმოადგენს ჩამოტვირთული მასალის გასუფთავება. ჩამოტვირთული მასალი ლინგვისტური დამუშავება სპეციალური კორპუსული ინსტრუმენტის – BootCatTools-ის საშუალებით განხორციელდა. ვებრესურსების ლინგვისტური დამუშავებისათვის ასევე ვეყრდნობით, ე.წ. „კლინივალის“ (CleanEval: cleaning webpages) გამოცდილებას (Baroni, Chantree, Kilgarrieff, Sharoff 2008). „კლინივალი“ წარმოადგენს ვებგვერდის შეფასებას და მიზნად ისახავს ვებმონაცემების კორპუსულ რესურსად გამოყენებას. უნდა აღინიშნოს, რომ ჩვენ მიერ ვებგვერდების გასუფთავებაში გამოყენებული მეთოდოლოგია არის უახლესი და არის ლიდის უნივერსიტეტის (დიდი

ბრიტანეთი), ტრენტოს უნივერსიტეტისა (იტალია) და „ლექსიკალ კომპიუტინგის“ (Lexical Computing LTD) ერთობლივი პროექტი.

**სურათი 3: ქართული ენის ფუნქციური სიტყვების შერჩევის პროცესი**

	A	B
1	Word	Meaning
2	და	And / Sister
3	რომ	that (conjunction)
4	ასევე	as well as
5	ეს	this
6	ამ	this
7	იყო	was
8	მაგრამ	but
9	მის	her/his/its
10	რა	what
11	არის	is
12	ის	that (conjunction), he/she/it
13	როგორც	as
14	რომელიც	which/that
15	რაც	That (conjunction)
16	მისი	her/his/its
17	თუმცა	but
18	ძალიან	very
19	შორის	(between, among)
20	მიერ	by (sb/smth)

დღეს ძალიან ბევრი იყენებს ვებს, როგორც მონაცემთა წყაროს კორპუსისათვის (Baroni & Bernardini 2006; Fairon 2007, Hundt 2007, Kilgarriff & Grefenstette 2003), რაც თავის მხრივ, გარკვეულ პრობლემებთან და სირთულეებთანაა დაკავშირებული. ვებგვერდების ჩამოტვირთვისას თავს იჩენს შემდეგი საკითხები:

1. როგორ უნდა გამოირიცხოს ლინგვისტურად არარელევანტური ინფორმაცია (“boilerplate”), მაგალითად, როგორცაა სპეციალური ნიშნები და სიმბოლოები, კოლონტიტულები (headers, footers) და სხვ.;
2. როგორ უნდა განისაზღვროს პარაგრაფები და სხვა სტრუქტურული ხასიათის ინფორმაცია;
3. როგორ უნდა მივიღოთ სტანდარტული ფორმატის ტექსტი, რომელიც გამოდგება შემდგომი ლინგვისტური დამუშავებისათვის.

ზემოთ მოცემული საკითხები ძალიან მნიშვნელოვანი, ვინაიდან, რაც უფრო კარგად არის გათვალისწინებული აღნიშნული საკითხები, მით უფრო კარგი შედეგის მიღწევაა შესაძლებელი. ლინგვისტური დამუშავების შემდგომი დონეები სწორედ ჩამოტვირთული მასალის „სისუფთავეზე“ დამოკიდებული. იმ შემთხვევაში, თუ ლინგვისტურ დამუშავებამდე ვებგვერდიდან ჩამოტვირთული მასალა არაა „გასუფთავებული“, ყველაზე ხშირად განმეორდება ისეთი ბიგრამები, როგორიცაა „დააჭირეთ აქ“ (click here) და „ვრცლად“/ „დამატებითი ინფორმაციისათვის“ / „იხილეთ“ (Further information). რასაკვირველია, ყოველივე ეს მნიშვნელოვნად შეცვლის ენის რეალურ სურათს.

როგორც უკვე აღინიშნა, ქართული ენის ვებკორპუსის შექმნის პირველ ეტაპს სწორედ მასალის შერჩევა (data selection) წარმოადგენდა - შესაბამისი ვებგვერდების შერჩევა და მისი მომზადება ჩამოსატვირთად. აღსანიშნავია, რომ ქართული ენის ვებკორპუსისათვის მხოლოდ HTML ფორმატის ვებგვერდებია გამოყენებული. PDF-ისა და ვორდის (Word) ფაილების ჩამოტვირთვა სხვა სახის პრობლემებთანაა დაკავშირებული, მაგალითად, როგორცაა კოდირება, ტექსტში გამოყენებული სპეციფიკური შრიფტებისა და სიმბოლოების შემდგომი დამუშავება და ა.შ. ამ ეტაპზე აღნიშნული ფორმატის ვებრესურები არ არის მოცემული კორპუსში. თუმცა, აქვე უნდა აღინიშნოს, რომ ხელით დამუშავდა სპეციალურად შერჩეული სამეცნიერო ხასიათის ტექსტები (PDF, Word), სულ 235 ფაილი და გაერთიანდა კორპუსში.

ქართული ენის ვებკორპუსის ვებგვერდების ანოტირება და გასუფთავება მოიცავდა შემდეგ პუნქტებს:

1. HTML-ისა და ჯავასკრიპტების (JavaScript) კოდების მოხსნას, ლინგვისტურად არარელევანტური ინფორმაციის წაშლას (კოლონტიტულები, საავტორო უფლების აღმნიშვნელი ნიშნები, ბმულების, ცალკეულ ვებგვერდებში გამეორებული მასალები და ა.შ.);
2. მინიმალური სიმბოლოების გამოყენებით ძირითადი კოდების მინიჭება (იხილეთ სურათი 3) ტექსტების სათაურებისათვის, პარაგრაფებისათვის და ა.შ.





#### 14 რეპრეზენტაციულობა ქართული ენის ვებკორპუსში

ნებისმიერი ტიპის კორპუსის ერთ-ერთ აუცილებელ კრიტერიუმს რეპრეზენტაციულობა წარმოადგენს. თუმცა, კორპუსის ლინგვისტებს შორის არ არსებობს საერთო აზრი იმის თაობაზე, თუ რა არის მიჩნეული რეპრეზენტაციულად და ზოგადად, როგორ შეიძლება რეპრეზენტაციულობის მიღწევა. გამომდინარე ზემოაღნიშნულიდან, ტერმინი „რეპრეზენტაციულობა“ ნაკლებად ან საერთოდ არ გამოიყენება. ამ უკანასკნელის ნაცვლად დღეს ფართოდ გამოიყენება ტერმინი „დაბალანსება“ / „დაბალანსებული“ (balance / balanced).

ქართულ ენის ვებკორპუსის შექმნის საწყის ეტაპზე მკაცრად განისაზღვრა შემდეგი:

1. კორპუსის ტიპი
2. შეირჩა შესაბამისი ვებსაიტები.

ვინაიდან სადოქტორო კვლევის მიზანს თანამედროვე ქართული ენის სისწორული ლექსიკის განსაზღვრა წარმოადგენდა, ქართული ენის ვებკორპუსი ასახავს თანამედროვე ქართულ ბუნებრივ ენას, კერძოდ, ენის განვითარების ბოლო 10-15 წელიწადს. ვებკორპუსის ტექსტური მასალის ჩამოტვირთვა განხორციელდა სახელმწიფო, საგანმანათლებლო, არასამთავრობო, კერძო ორგანიზაციების ვებსაიტებიდან და რაც ყველაზე მნიშვნელოვანია, პირადი ბლოგებიდან და ფორუმებიდან. იმისათვის, რომ კორპუსში მაქსიმალურად ასახულიყო ბუნებრივი, ცოცხალი, თანამედროვე ქართული ენა, თავიდანვე გამოირიცხა პოეზია და თარგმანები. ამგვარად, კორპუსში არ არის შესული პოეზია და თარგმანები. როგორც ვიცით, პოეზიისათვის და თარგმანებისათვის დამახასიათებელია სპეციფიკურად შერჩეული ლექსიკა, რომელიც რამდენადმე შორს დგას ბუნებრივი ენისგან. ქართული ენის ვებკორპუსისათვის წინასწარ განისაზღვრა და შეირჩა **697 ვებსაიტი** (იხილეთ სურათი 4), რომლებმაც საბოლოო ჯამში შეადგინა 618468 ვებგვერდი. აღნიშნული ვებგვერდებიდან ჩამოტვირთული მასალა კი დაახლოებით მოიცავს 150 მილიონ სიტყვას.



## სურათი 5: KaWaC-ის ვებსაიტების შერჩევა

A	B
1	<b>MINISTRIES</b>
2	<b>Ministry of Economy and Sustainable Development of Georgia</b> <a href="http://www.economy.ge/">http://www.economy.ge/</a>
3	<b>Ministry of Science and Education of Georgia</b> <a href="http://www.mes.gov.ge/">http://www.mes.gov.ge/</a>
4	<b>Ministry of Environment and Natural Resources Protection of Georgia</b> <a href="http://moe.gov.ge/">http://moe.gov.ge/</a>
5	<b>Ministry of Energy of Georgia</b> <a href="http://www.energy.gov.ge/">http://www.energy.gov.ge/</a>
6	<b>Ministry of Defense of Georgia</b> <a href="http://www.mod.gov.ge/">http://www.mod.gov.ge/</a>
7	<b>Ministry of Justice of Georgia</b> <a href="http://www.justice.gov.ge/">http://www.justice.gov.ge/</a>
8	<b>Ministry of Culture and Monument Protection of Georgia</b> <a href="http://www.culture.gov.ge/">http://www.culture.gov.ge/</a>
9	<b>Ministry of Internally Displaced Persons from the Occupied Territories, Accomodation and Refugees of Georgia</b> <a href="http://www.mra.gov.ge/">http://www.mra.gov.ge/</a>
10	<b>Ministry of Regional Development and Infrastructure of Georgia</b> <a href="http://www.mrdi.gov.ge/ge/main">http://www.mrdi.gov.ge/ge/main</a>
11	<b>Ministry of Foreign Affairs of Georgia</b> <a href="http://www.mfa.gov.ge/">http://www.mfa.gov.ge/</a>
12	<b>Ministry of Corrections and Legal Assistance of Georgia</b> <a href="http://www.mcla.gov.ge/">http://www.mcla.gov.ge/</a>
13	<b>Ministry of Agriculture of Georgia</b> <a href="http://moa.gov.ge/">http://moa.gov.ge/</a>
14	<b>Ministry of Sport and Youth Affairs of Georgia</b> <a href="http://msy.gov.ge/">http://msy.gov.ge/</a>
15	<b>Ministry of Finance of Georgia</b> <a href="http://www.mof.gov/">http://www.mof.gov/</a>
16	<b>Ministry of Internal Affairs of Georgia</b> <a href="http://police.ge/">http://police.ge/</a>
17	<b>Ministry of Labor Health and Social Affairs of Georgia</b> <a href="http://www.moh.gov.ge/">http://www.moh.gov.ge/</a>
18	<b>The Government of Georgia</b> <a href="http://www.government.gov.ge/">http://www.government.gov.ge/</a>
19	<b>The Parliament of Georgia</b> <a href="http://www.parliament.ge/">http://www.parliament.ge/</a>
20	<b>Administration of President of Georgia</b> <a href="https://president.gov.ge/ge/President/PresidentialPow">https://president.gov.ge/ge/President/PresidentialPow</a>
21	
22	
23	<b>Library</b>
24	<b>National Parliamentary Library of Georgia</b> <a href="http://www.nplg.gov.ge/">http://www.nplg.gov.ge/</a>
25	
26	<b>Universities</b>
27	<b>Ivane Javakhishvili Tbilisi State University</b> <a href="http://www.tsu.edu.ge/">http://www.tsu.edu.ge/</a>
28	<b>Tbilisi State Medical University</b> <a href="http://www.tsmu.edu/tsmu12/tsmu/">http://www.tsmu.edu/tsmu12/tsmu/</a>
29	<b>Ilia State University (ISU)</b> <a href="http://www.iliauni.edu.ge/">http://www.iliauni.edu.ge/</a>
30	<b>Kutaisi Akaki Tsereteli State University</b> <a href="http://www.atsu.edu.ge/">http://www.atsu.edu.ge/</a>
31	<b>The Shota Rustaveli Theatre and Film State University</b> <a href="http://www.taflu.edu.ge/">http://www.taflu.edu.ge/</a>
32	<b>Batumi Shota Rustaveli State University</b> <a href="http://www.bsu.edu.ge/">http://www.bsu.edu.ge/</a>

მე-4 სურათზე ნაჩვენებია ქართული ენის ვებკორპუსისათვის ვებსაიტების განსაზღვრის და შერჩევის პროცესი. როგორც სურათიდან ჩანს, ვებკორპუსისათვის ძალიან დიდი სიფრთხილით შეირჩა თითოეული ვებსაიტი, ე.ი. ის ტექსტური მასალა, რაც კორპუსში უნდა შესულიყო.

ქართული ენის კორპუსი მოიცავს ტექსტურ მასალას შემდეგი ვებწყაროებიდან:

- 1) სამინისტროები და სხვა ოფიციალური ვებსაიტები - **65 ვებსაიტი**;
- 2) უნივერსიტეტები, ინსტიტუტები, კოლეჯები, პროფესიული სასწავლებლები - **64 ვებსაიტი**;
- 3) პირადი ბლოგები და ფორუმები - **417 ვებსაიტი**;
- 4) პრესა და ახალი ამბები - **124 ვებსაიტი**;
- 5) არასმთავრობო ორგანიზაციები, ბანკები, კერძო კომპანიები - **72 ვებსაიტი**;
- 6) მხატვრული ლიტერატურა (პროზა) - **235 თანამედროვე ქართველი ავტორის შემოქმედება**;
- 7) სამეცნიერო ნაშრომები/სტატიები, სადოქტორო /სამაგისტრო ნაშრომები: **246 ნაშრომი** და ა.შ.

## 2. ჟანრული კლასიფიკაცია ქართული ენის ვებკორპუსში

როგორც უკვე აღინიშნა, კორპუსულ ლინგვისტიკაში, კერძოდ, ენის კორპუსების მიმართ ნაკლებად გამოიყენება ტერმინი „რეპრეზენტაციულობა“, ვინაიდან ძნელია იმის თქმა, თუ რისი რეპრეზენტაციული შეიძლება იყოს კორპუსი. ამიტომ, აღნიშნული ტერმინის ნაცვლად გამოიყენება „დაბალანსება“ / „დაბალანსებული“ (balanced). შესაბამისად, შეიცვალა მიდგომაც, კორპუსის მიზანი არ არის იყოს რეპრეზენტაციული, არამედ შეიძლება იყოს დაბალანსებული ამა თუ იმ კონკრეტული საკვლევი მიზნებიდან გამომდინარე.

რამ განაპირობა აღნიშნული ცვლილება? პასუხი მარტივია, კორპუსის ლინგვისტიკის განვითარებამ და შესაბამისად, ახალი ტიპის კორპუსების გაჩენამ, მაგალითად, როგორცაა ვებკორპუსები. ვებკორპუსების მიმართ მართებულია ტერმინი „დაბალანსებულის“ გამოყენება, ვიდრე „რეპრეზენტაციულის“, ვინაიდან ვებკორპუსი შეუძლებელია იყოს რაღაცის რეპრეზენტაციული რამის მიმართ. ამგვარად, ვებკორპუსებში დაბალანსება განსხვავდება ტრადიციული ტიპის კორპუსების დაბალანსებისაგან. მაგალითად, ბრაუნის (the Brown University Standard Corpus of Present-Day American English), LOB-ისა (Lancaster-Oslo-Bergen Corpus) და LCMC-ის (the Lancaster Corpus of Mandarin Chinese) კორპუსში მოცემულია 15 ჟანრული კატეგორია, მაგალითად, ა) პრესა: რეპორტაჟი, ბ) პრესა: საგამომცემლო, გ) პრესა: მიმოხილვა, დ) რელიგია, ე) უნარები და ჰობი და ა.შ.

რაც შეეხება ბრიტანულ ნაციონალურ კორპუსს, მასში მოცემულია 4,000 ტექსტი, რომელიც 70 ჟანრულ კატეგორიად არის დაჯგუფებული, მაგ., მედიცინა, ტექნიკა, პრესა და ა.შ. არსებობს უფრო დეტალური და ვრცელი ჟანრული კლასიფიკაცია, მაგალითად, ადამციკთან (Adamzik 1995) 4000 ჟანრული კლასიფიკაციაა მოცემული. იხილეთ სურათი 6.

## სურათი 6: ადამციკის უანრული კლასიფიკაცია

Traditional research on genre classification Measuring text similarity	Need for genre categories Reliability problems	
<b>List from Adamzik (1995)</b>		
<i>Abänderungsantrag</i> <i>Abbestellung</i> <i>Abbruchgenehmigung</i> <i>Abdankungserklärung</i> <i>Abecedarium</i> <i>Abendblatt</i> <i>Abendgebet</i> <i>Abendgespräch</i> <i>Abendnachrichten</i> <i>Abendprogramm</i> <i>Abendzeitung</i> <i>Abenteuerroman</i> <i>Aberkennung</i> <i>Abfahrtsplan</i> <i>Abfindungserklärung</i> <i>Abgabebewilligung</i> <i>Abgabeordnung</i> <i>Abgangsmeldung</i> <i>Abgangszeugnis</i> <i>Abgeordnetenrede</i> <i>Abgesang [im Meistersang]</i> <i>Abhandlung</i> <i>Abhang [ind. Hymne]</i> <i>Abhörverbot</i> <i>Abiturientenzeugnis</i> <i>Abiturzeugnis</i> <i>Abkommen</i>	<i>Abrüstungsverhandlungen</i> <i>Absage</i> <i>Absatz</i> <i>Absatzgarantie</i> <i>Abschiedsbrief</i> <i>Abschiedsgespräch</i> <i>Abschiedsrede</i> <i>Abschilderung</i> <i>Abschlußarbeit</i> <i>Abschlußbesprechung</i> <i>Abschlußbilanz</i> <i>Abschlußgespräch</i> <i>Abschlußrechnung</i> <i>Abschlußzeugnis</i> <i>Abschnitt</i> <i>Abschrift</i> <i>Abschußliste</i> <i>Abschußplan</i> <i>Abschwörungsformel</i> <i>Absichtserklärung</i> <i>Absolutorium [Reifezeugnis; österr.: Bestätigung einer Hochschule über erbrachte Leistungen]</i> <i>Abstammungsklage</i> <i>Abstammungsnachweis</i> <i>Abstammungsurkunde</i> <i>Abstimmungsunterlagen</i>	<i>Adversaria [vor Augen liegende Kladde mit ungeordneten Konzepten, Notizen]</i> <i>Agenda [Notizbuch]</i> <i>Agende [Kirch]</i> <i>Agentenroman</i> <i>Agenturbericht</i> <i>Agenturmeldung</i> <i>Agitpropstück</i> <i>Ahnenprobe</i> <i>Ahmentafel</i> <i>Akkordzettel</i> <i>Akkreditiv [Beglaubigungsschreiben eines Diplomaten]</i> <i>Akquisitionsliste [Anschaffungsliste]</i> <i>Akte</i> <i>Aktenband</i> <i>Aktenfaszikel</i> <i>Aktenheft</i> <i>Aktennotiz</i> <i>Aktenstück</i> <i>Aktenvermerk</i> <i>Aktie</i> <i>Aktiengesetz</i> <i>Akzept [Bank]</i> <i>Akzessionsliste [Verzeichnis von</i>
Serge Sharoff	Reliable genre classification	



როგორც ზემოთ აღინიშნა, ვებკორპუსების უანრული კლასიფიკაციები განსხვავდება ტრადიციული კორპუსებისთვის შემუშავებული კლასიფიკაციისაგან, რადგან ინტერნეტმა და ვებრესურსებმა განაპირობა ისეთი უანრების წარმოშობა, რომლებიც ტრადიციულ კორპუსებში არ არის წარმოდგენილი. ამგვარად, გავითვალისწინეთ ვებკორპუსის უანრული კლასიფიკაციის უახლესი მეთოდები და მიდგომები, კერძოდ, მელერის, შაროვისა და სანტინის (Mehler, Sharoff, Santini 2010) ვების უანრული კლასიფიკაციები. იხილეთ სურათი 7.

სურათი 7: ვებჟანრების მაგალითები: მელერი, შაროფი, სანტინი (2010)

Traditional research on genre classification Measuring text similarity		Need for genre categories Reliability problems	
Examples of webgenres			
Mehler, Sharoff, Santini, (2010). Genres on the Web.			
MGC (20 genres)		KI-04 (8 genres)	Santinis (8 genres)
adults	informative	article	blog
blog	journalistic	discussion	<b>eShop</b>
childrens	official	download	<b>faq</b>
commercial	personal	<b>help</b>	frontpage
community	poetry	linklists	hotlist
content delivery	prose	portrait-nonpriv	homepage
entertainment	scientific	portrait-priv	sitemap
error message	<b>shopping</b>	<b>shop</b>	Search page
<b>FAQ</b>		user input	
gateway			
index			

ქართული ენის ვებკორპუსის ტექსტების ანოტირებისათვის გამოვიყენეთ შაროფისა და ფორსის (Forsyth & Sharoff, 2014) ფუნქციური ჟანრული განზომილებები (Functional Genre Dimensions), რომლებიც მოიცავს 17 ჟანრულ კატეგორიას და კითხვარების სახით არის შედგენილი და თითოეულ კითხვარს შესაბამისი საიდენტიფიკაციო კოდი აქვს მინიჭებული. იხილეთ სურათი 8 ფუნქციური ჟანრული განზომილებებისათვის შესაბამისი კითხვარებითა და კოდებითურთ.

## სურათი 8: ფუნქციური ჟანრული განზომილებები

Code	Question to be answered
A1.	To what extent does the text argue to persuade the reader to support (or renounce) an opinion or a point of view? ( <i>Strongly</i> , if argumentation is obvious)
A3.	To what extent is the text concerned with expressing feelings or emotions? ( <i>None</i> for neutral explanations, descriptions and/or reportage.)
A4.	To what extent is the text's content fictional? ( <i>None</i> if you judge it to be factual/informative.)
A5.	To what extent is the text light-hearted, i.e. aimed mainly at amusing or entertaining the reader? ( <i>None</i> if it appears earnest or serious; even when it tries to keep the reader interested and involved)
A6.	To what extent is the text written in an informal style, using colloquialism and/or slang (as opposed to the "standard" or "prestige" variety of language)?
A7.	To what extent does the text aim at teaching the reader how to do something? (e.g. a tutorial)
A8.	To what extent does the text appear to be a news report such as might be found in a newspaper, i.e. an informative report of recent events? (recent at the time of writing. <i>None</i> if a news source does not provide new information about what happened, while analysing information from other sources).
A9.	To what extent does the text lay down a contract or specify a set of regulations? (This includes copyright notices.)
A11.	To what extent does the text report a first-person point of view?
A12.	To what extent does the text promote a commercial product or service?
A13.	To what extent is the text intended to promote a political movement, party, religious faith or other non-commercial cause? (i.e. any promotion of not-for-profit causes)
A14.	To what extent does the text belong to the field of Science, Technology and/or Engineering? (As opposed to the Arts, Humanities &/or Social Studies. This is not necessarily a research paper. A newswire text can include scientific contents, so it can be judged as <i>Strongly</i> or <i>Partly</i> .)
A15.	To what extent does the text require background knowledge of a specialist or access to a reference source in order to be comprehensible for the so-called "general reader"?
A16.	To what extent does the text provide information to define a topic? (For example, <i>Strongly</i> for encyclopedic articles or text books).
A17.	To what extent do you judge the text to evaluate something? (For example, by providing a review).
A18.	To what extent does the text contain active interaction between several participants? (For example, forums or dialogue lines coming from theatre plays).
A19.	To what extent does the author of the text pay attention to aesthetic appearance of its language? ( <i>Strongly</i> for poetry, language experiments, etc.)

ფუნქციურ ჟანრულ განზომილებებს გააჩნია შეფასების შემდეგი სქემა:

- 0 ქულა - არა, საერთოდ არ განეკუთვნება ამ კატეგორიას (non or hardly at all);
- 0.5 ქულა - ოდნავ განეკუთვნება (slightly);
- 1 ქულა - ნაწილობრივ განეკუთვნება (somewhat or partly);
- 2 ქულა - განეკუთვნება (strongly or very much so).

ქართული ენის ვებკორპუსის ტექსტების ანოტირება სწორედ ფუნქციური ჟანრული კატეგორიების მიხედვით წარიმართა და კორპუსში მოცემული 618468 ვებგვერდიდან შეირჩა 1067 ვებგვერდი ანოტირებისათვის. იხილეთ სურათი 9.

**სურათი 9: ფუნქციური ჟანრული განზომილებების მიხედვით  
ვებგვერდების ანოტირება**

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U
1	URL	ID	A1	A3	A4	A5	A6	A7	A8	A9	A11	A12	A13	A14	A15	A16	A17	A18	A19	Text summary	
2	1234friends.wordpress.com/page/2/	1	0	2	0.5	1	1	0	0	0	2	0	0	0	0	0	0	0	0	6.5	
3	1234friends.wordpress.com/2013/05/25/%83%99%	2	0	0	0	2	0	0	0	0	2	0	0	0	0	0	0	0	0	4	
4	150ekvtime.wordpress.com/%83%91%83%98%	3	0	0	0	2	0	0	0	0	0	0	0	0	0	2	0	0	0	4	
5	150ekvtime.wordpress.com/category/%83%94%	4	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	2	
6	4bloging.wordpress.com/%83%83%94%83%83%	5	0	0	0	0	0	1	0	2	0	0	0	0	0	0	0	0	0	3	
7	4bloging.wordpress.com/2012/05/27/%83%83%	6	0	0	0	2	0	0	0	0	0	0	0	0	0	2	0	0	0	4	
8	4bloging.wordpress.com/page/2/	7	0	1	1	2	0.5	0	0	0	0.5	0	0	0	0	0	0	0	0	5	
9	abadu.wordpress.com/about/	8	0	2	0	2	1	0	0	0	2	0	0	0	0	0	0	1	0	8	
10	abadu.wordpress.com/category/%83%90%83%	9	0	2	0	2	0	0	0	0	2	0	0	0	0	0	0	0	0	6	
11	abadu.wordpress.com/category/%83%98%83%	10	0	2	0	2	0	0	0	0	2	0	0	0	0	0	0	0	0	6	
12	aboutguria.blogspot.co.uk/2013/10/blog-post_4.1	11	2	1	0	1	0	0	0	0	0.5	0	0	0	0	1	1	0.5	0	7	
13	aboutguria.blogspot.co.uk/2013/04/blog-post_22	12	1	1	0	2	0.5	0	0	0	2	0	0	0	0	0	1	0	0	7.5	
14	aboutguria.blogspot.co.uk/2012/05/blog-post_18	13	1	1	0	0.5	0	0	0	0	2	0	0	0	0	0	0	0	0	4.5	
15	active.ge/category/%83%83%98%83%9A%83%	14	0	0	0	0	0	0.5	0	0	0	2	0	0	0	0	0	0	0	2.5	
16	active.ge/category/%83%83%94%83%83%83%	15	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	2	
17	adolis.wordpress.com/2013/02/07/agraru-li-unive	16	0	2	0	2	0.5	0	0	0	2	0	0	0	0	0	0	2	0	8.5	
18	adolis.wordpress.com/2012/06/27/%83%83%9C	17	0	2	0	2	0	0	0	0	2	0	0	0	0	0	0	0	0	6	
19	afba.ge/main/71--.html	18	2	0	0	0	0	0	2	0	0	0	0	0	0.5	0	2	0	0	6.5	
20	afba.ge/main/68--.html	19	2	0	0	0	0	0	2	0	0	0	0	0	0.5	0	2	0	0	6.5	
21	afba.ge/main/64--.html	20	2	0	0	0	0	0	2	0	0	0	0	0	0.5	0	2	0	0	6.5	
22	akakisanadze.com/author/lps1520/	21	0	0	0		1	2	0	0	2	0	0	0	0	0	0	0	0	5	
23	akakisanadze.com/category/websites-%83%95%	22	0	0	0	0	1	2	0	0	2	0	0	0	0	0	0	0	0	5	
24	akunkula.wordpress.com/%83%9B%83%94/	23	0	2	0	2	0	0	0	0	2	0	0	0	0	0	0	0	0	6	
25	akunkula.wordpress.com/2013/12/28/my_er/	24	0	2	0	2	1	0	0	0	2	0	0	0	0	0	0	1	0	8	
26	all4people.wordpress.com/category/%83%83%83%	25	0	0	0	2	0	0	0	0	0	0	0	0	0	1	0	0	0	3	
27	all4people.wordpress.com/category/%83%83%83%	26	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0	0	0	2	
28	all4people.wordpress.com/page/2/	27	0	2	0	2	2	0	0	0	2	0	0	0	0	0	0	2	0	10	
29	alternativa.ge/2011/05/23/109/	28	0	2	0	2	1	0	0	0	2	0	0	0	0	0	0	0	0	7	

ამგვარად, ქართული ენის ვებკორპუსში ჟანრული კლასიფიკაცია გარჩეულია ფუნქციური ჟანრული განზომილებების მიხედვით. შერჩეული 1067 ვებგვერდიდან მეტ-ნაკლებად ხშირად ხმარებული შემდეგი ჟანრული კლასებია: 3 - პირადი ემოციების გამომხატველი შინაარსის ვებგვერდები (პირადი ბლოგები, ფორუმები), 4 - მხატვრული ლიტერატურა (პროზა), 7 - ინსტრუქციული (კულინარია, მედიცინა და ა.შ.), 8 - პრესა: ახალი ამბები.



### 3. მულტიტექსტური მორფოსინტაქსური პარამეტრები

კორპუსის მასალის ამა თუ იმ კონკრეტული ლინგვისტური კვლევებისათვის გამოყენებისათვის, როგორცაა მაგალითად სიხშირული ლექსიკის დადგენა და ა.შ. აუცილებელია კორპუსის მორფოლოგიური ანოტირება.

ქართული ენის ვებკორპუსის მორფოლოგიური ანოტირების პლატფორმად გამოყენებულია „მულტიტექსტური მორფოსინტაქსური პარამეტრები“ (MULTEXT-East Morphosyntactic Specifications - MTE), მეოთხე ვერსია.

(<http://nl.ijs.si/ME/V4/msd/html/index.html>). „მულტიტექსტური მორფოსინტაქსური პარამეტრები“ არის მრავალენოვანი მონაცემთა ბაზა ენის ლინგვისტური დამუშავებისათვის, კერძოდ, ორიენტირებულია მორფოსინტაქსურ ანოტირებაზე. ამ ბაზის რესურსები მოიცავს ცენტრალურ და აღმოსავლეთ ევროპულ ენებსა და EAGLES-ზე (Expert Advisory Group on Language Engineering Standards) დაფუძნებულ პარამეტრებს, მორფოსინტაქსურ ლექსიკონს, ანოტირებულ პარალელურ, შედარებით (comparable) და მეტყველების (speech corpus) კორპუსებს. „მულტიტექსტური მორფოსინტაქსური პარამეტრების“ მეოთხე ვერსიაში წარმოდგენილია XML-ში კოდირებული მორფოსინტაქსური პარამეტრები 16 ენისათვის: ბულგარული, ხორვატული, ჩეხური, ინგლისური, ესტონური, უნგრული, ლიტვური, მაკედონური, ირანული, პოლონური, რუსული, სერბიული, სლოვაკური, სლოვენური, უკრაინული და სხვ. პროექტი „მრავალენოვანი ტექსტური ინსტრუმენტები და კორპუსები აღმოსავლეთ და ცენტრალური ევროპული ენებისათვის“ (Multilingual Text Tools and Corpora for Eastern and Central European Languages) ხორციელდება 1995/97 წლიდან.

„მულტიტექსტური მორფოსინტაქსური პარამეტრების“ მეოთხე ვერსია შეესაბამება ტექსტის კოდირების/ანოტირების ინიციატივას (Text Encoding Initiative) – TEI P5-ს. „მულტიტექსტური მორფოსინტაქსური პარამეტრები“ მოიცავს მოცემული ენის მეტყველების ნაწილების კატეგორიებს, რომელთაც ძირითადი კატეგორიები ეწოდებათ (main categories). ძირითად კატეგორიები დაჯგუფებულია მეტყველების ნაწილების თავისებურებების მიხედვით და მინიჭებული აქვს შესაბამისი კოდები, რასაც მორფოსინტაქსური აღწერა (MSD – morphosyntactic Description) ეწოდება.

„მულტიტექსტური მორფოსინტაქსური პარამეტრები“ განსაზღვრავს ძირითად მორფოლოგიურ კატეგორიებს და მათთან დაკავშირებულ ქვეკატეგორიებს

წყვილებს, ე.ი. „მორფოსინტაქსურ აღწერას“ - (Morphosyntactic Description (MSD)). მაგალითად:

**არსებითი სახელი, ტიპი=საზოგადო, რიცხვი=მხოლოდითი, ბრუნვა=სახელობითი**

ზემოთ წარმოდგენილი ქვეკატეგორიების წყვილების სტრუქტურას შეესაბამება „მორფოსინტაქსური აღწერის“ (ე.ი. MSD) ტაგი/მარკერი: Ncsn (*Noun, Type=Common, Number=singular, Case=Nominative*), რაც გამოყენებულია მორფოლოგიური ანოტირებისას (Sharoff, et al, 2008).

ამგვარად, ქართული ენისათვის მეტყველების ნაწილების მარკერების სისტემების, ე.წ. ტაგსეტების (tagsets) შემუშავება და მორფოლოგიური ანოტირება „მულტიტექსტური მორფოსინტაქსური პარამეტრების“ მეოთხე ვერსიის შესაბამისად განხორციელდა. ეს უკანასკნელი ძალზე მნიშვნელოვანია იმ თვალსაზრისითაც, რომ ქართული ენის ვებკორპუსი შეიძლება შევადაროთ მორფოსინტაქსური პარამეტრების ბაზაში მოცემული 16 ენის კორპუსებს.

### **3.1 მეტყველების ნაწილების მარკერების სისტემა ქართული ენისათვის**

ქართული ენის ტაგსეტები, ერთი მხრივ, დაეფუძნა „მულტიტექსტური მორფოსინტაქსური პარამეტრების“ მეოთხე ვერსიას, ხოლო, მეორე მხრივ, კი აკაკი შანიძის ქართული ენის გრამატიკის საფუძვლებს (შანიძე 1973).

ქართული ენის ტაგსეტების შემუშავების პროცესში გავითვალისწინეთ ქართული ენის მორფოლოგიის სირთულე, შესაბამისად, შემუშავდა ზოგადი სტრატეგია, რომელიც მოიცავს ქართული ენის მორფოლოგიის ყველაზე მნიშვნელოვან ასპექტებს (იხილეთ ცხრილი 1). ტაგსეტებში მაქსიმალურად გამოყენებულია მორფოსინტაქსური პარამეტრების კოდები. გამონაკლისს წარმოადგენს მხოლოდ ქართული ენის მორფოლოგიისათვის დამახასიათებელი თავისებურებანი, რომელთაც მორფოსინტაქსური პარამეტრები არ მოიცავს, ამ შემთხვევაში შემუშავდა ახალი კოდები ქართული ენის მორფოლოგიური თავისებურებების ასახვის მიზნით.



**ცხრილი 1: მეტყველების ნაწილების მარკირების სისტემა:  
არსებითი სახელი**

P	Attribute	Value	Attribute	Georgian
0	CATEGORY	Noun	N	არსებითი სახელი
1	Type	Common	C	საზოგადო
		Proper	P	საკუთარი
2	Number	Singular	S	მხოლოდითი
		Plural	P	მრავლობითი
3	Case	nominative	N	სახელობითი
		Ergative	E	მოთხრობითი
		Dative	D	მიცემითი
		Genitive	G	ნათესაობითი
		instrumental	I	მოქმედებითი
		Adverbial	R	ვითარებითი
		Vocative	V	წოდებითი
4	Particles	-c	C	-ც (ნაწილაკი)
		-gha	gha	-ღა (ნაწილაკი)
		-ve	ve	-ვე (ნაწილაკი)
5	Adposition	-vit	V	-ვით (თანდებულიანი)
		tan	T	-თან (თანდებულიანი)
		-ze	Z	-ზე (თანდებულიანი)
		-shi	H	-ში (თანდებულიანი)
		-tvis	S	-თვის (თანდებულიანი)
		-gan	G	-გან (თანდებულიანი)
		-ken	K	-კენ (თანდებულიანი)
		-ebr	B	-ებრ (თანდებულიანი)
		-tanave	N	-თანავე (თანდებულიანი)
		-gamo	O	-გამო (თანდებულიანი)
		-dami	A	-დამი (თანდებულიანი)
		-dmi	m	-დმი (თანდებულიანი)
		-urt	U	-ურთ (თანდებულიანი)
		-idan	I	-იდან (თანდებულიანი)
		-dan	D	-დან (თანდებულიანი)
		-mde	E	-მდე (თანდებულიანი)

ცხრილი 1.1: მეტყველების ნაწილების მარკირების სისტემა: ზმნა

P	Attribute	Value	Code	Georgian
0	<b>CATEGORY</b>	<b>Verb</b>	<b>V</b>	<b>ზმნა</b>
	VForm	Indicative	i	ძირითადი ფორმა
		imperative	m	ბრძანებითი
1	Person	first	1	I პირი
		second	2	II პირი
		third	3	III პირი
2	Number	singular	s	მხოლოდითი
		plural	p	მრავლობითი
3	Tense/Series	present	p	აწმყო
		future	f	მყოფადი
		imperfect	i	უწყვეტელი
		conditional	c	ხოლმეობითი
		present subjunctive	w	აწმყოს კავშირებითი
		future subjunctive	y	მყოფადის კავშირებითი
		aorist	a	წვეტილი
		Optative (II subjunctive)	o	II კავშირებითი
		perfect (first evidential)	e	I თურმეობითი
		pluperfect (second evidential)	l	II თურმეობითი
		III subjunctive (third evidential)	s	III კავშირებითი

**ცხრილი 12: მეტყველების ნაწილების მარკირების სისტემა:  
მასღარი**

P	Attribute	Value	Code	Georgian
<b>0</b>	<b>CATEGORY</b>	<b>Masdar</b>	<b>W</b>	<b>მასღარი</b>
1	Number	singular	s	მხოლოდობითი
		plural	p	მრავლობითი
2	Case	Nominative	n	სახელობითი
		Ergative	e	მოთხრობითი
		Dative	d	მიცემითი
		Genitive	g	ნათესაობითი
		Instrumental	i	მოქმედებითი
		Adverbial	r	ვითარებითი
		Vocative	v	წოდებითი

**ცხრილი 13: მეტყველების ნაწილების მარკირების სისტემა:  
მიმღეობა**

P	Attribute	Value	Code	Georgian
<b>0</b>	<b>CATEGORY</b>	<b>Participle</b>	<b>G</b>	<b>მიმღეობა</b>
1	Type	Past	g	ნამყოს მიმღეობა
		Future	f	მყოფადის მიმღეობა
2	Case	Nominative	n	სახელობითი
		Ergative	e	მოთხრობითი
		Dative	d	მიცემითი
		Genitive	g	ნათესაობითი
		Instrumental	i	მოქმედებითი
		Adverbial	r	ვითარებითი
		Vocative	v	წოდებითი

**ცხრილი 14: მეტყველების ნაწილების მარკირების სისტემა:  
შეღბენილი შემასმენელი**

P	Attribute	Value	Code	Georgian
0	CATEGORY	Compound Verb	Z	შეღბენილი შემასმენელი
1	Form	noun	n	არსებითი სახელი
		adjective	a	ზედსართავი სახელი
		adverb	r	ზმნიზედა
		adposition	s	თანდებული
		numeral	m	რიცხვითი სახელი
		pronoun	p	ნაცვალსახელი
		masdar	w	მასდარი
		participle	g	მიმღეობა
		particle	q	ნაწილაკი
		conjunction	c	კავშირი

**ცხრილი 15: მეტყველების ნაწილების მარკირების სისტემა:  
ზედსართავი სახელი**

P	Attribute	Value	Code	Georgian
0	CATEGORY	Adjective	A	zedsarTavi saxeli
1	Case	nominative	N	სახელობითი
		ergative	E	მოთხრობითი
		dative	D	მიცემითი
		genitive	G	ნათესაობითი
		instrumental	I	მოქმედებითი
		adverbial	R	ვითარებითი
		vocative	V	წოდებითი

**ცხრილი 1.6: მეტყველების ნაწილების მარკირების სისტემა:  
რიცხვითი სახელი**

<b>p</b>	<b>Attribute</b>	<b>Value</b>	<b>Code</b>	<b>Georgian</b>
0	<b>CATEGORY</b>	<b>Numeral</b>	<b>M</b>	<b>რიცხვითი სახელი</b>
1	Number	singular	S	მხოლოდითი
		plural	P	მრავლობითი
2	Case	nominative	N	სახელობითი
		ergative	E	მოთხრობითი
		dative	D	მიცემითი
		genitive	G	ნათესაობითი
		instrumental	I	მოქმედებითი
		adverbial	R	ვითარებითი
		vocative	V	წოდებითი

**ცხრილი 1.7: მეტყველების ნაწილების მარკირების სისტემა:  
ნაცვალსახელი**

<b>P</b>	<b>Attribute</b>	<b>Value</b>	<b>Code</b>	<b>Georgian</b>
0	<b>CATEGORY</b>	<b>Pronoun</b>	<b>P</b>	<b>ნაცვალსახელი</b>
1	Number	singular	S	მხოლოდითი
		plural	P	მრავლობითი
2	Case	nominative	N	სახელობითი
		ergative	E	მოთხრობითი
		dative	D	მიცემითი
		genitive	G	ნათესაობითი
		instrumental	I	მოქმედებითი
		adverbial	R	ვითარებითი
		vocative	V	წოდებითი

**ცხრილი 1.8: მეტყველების ნაწილების მარკირების სისტემა: ზმნიზება**

<b>P</b>	<b>Attribute</b>	<b>Value</b>	<b>Code</b>	<b>Georgian</b>
0	<b>CATEGORY</b>	<b>Adverb</b>	<b>R</b>	<b>ზმნიზება</b>

ცხრილი 1.9: მეტყველების ნაწილების მარკირების სისტემა: კავშირი

P	Attribute	Value	Code	Georgian
0	CATEGORY	Conjunction	C	კავშირი

ცხრილი 1.10: მეტყველების ნაწილების მარკირების სისტემა: შორისდებული

P	Attribute	Value	Code	Georgian
0	CATEGORY	Interjection	I	შორისდებული

ცხრილი 1.11: მეტყველების ნაწილების მარკირების სისტემა: ნაწილაკი

P	Attribute	Value	Code	Georgian
0	CATEGORY	Particle	Q	ნაწილაკი

ცხრილი 1.12: მეტყველების ნაწილების მარკირების სისტემა: თანდებული

P	Attribute	Value	Code	Georgian
0	CATEGORY	Adposition	S	თანდებული

ცხრილი 1.13: მეტყველების ნაწილების მარკირების სისტემა: აბრევიატურა

P	Attribute	Value	Code
0	CATEGORY	Abbreviation აბრევიატურა	Y
1	Number	singular	s
		plural	p
2	Case	nominative	n
		ergative	e
		dative	d
		genitive	g
		instrumental	i
		adverbial	r
		vocative	v

ცხრილი 1.14: მეტყველების ნაწილების მარკირების სისტემა: უცნობი

P	Attribute	Value	Code
0	CATEGORY	Residual/ უცნობი	X

ქართული ენის ტაგსეტები მოიცავს 15 ძირითად კატეგორიას, ესენია: არსებითი სახელი, ზმნა, ზედსართავი სახელი, ნაცვალსახელი, რიცხვითი სახელი, ზმნიზედა, თანდებული, კავშირი, ნაწილაკი, შორისდებული, მასდარი, მიმღეობა, შედგენილი შემასმენელი, აბრევიატურა და უცნობი (residual). თითოეული კატეგორიისათვის განსაზღვრულია შესაბამისი ატრიბუტული ჯგუფები (attribute value pairs), სულ 331 წყვილი.

უნდა აღინიშნოს, რომ ამ სისტემაში სიტყვა „კატეგორია“ არ უნდა ავურიოთ „გრამატიკულ კატეგორიაში“. „კატეგორია“ აქ არის “მულტიტექსტური მორფოსინტაქსური პარამეტრების” ტაგსეტების სისტემაში გამოყენებული ტერმინი, რომელიც მოცემულ სისტემაში ამა თუ იმ მეტყველების ნაწილის ჯგუფებად განლაგებას გულისხმობს. როგორც ზემოთ აღვნიშნე, ქართული ენის ტაგსეტებისათვის შევიმუშავეთ ზოგადი სტრატეგია, რომელიც მოიცავს ქართული ენის მორფოლოგიის ყველაზე მნიშვნელოვან ასპექტებს. ტაგსეტებში ცალკე ჯგუფად არ არის გამოტანილი ნართანიანი მრავლობითი. იგი მოცემულია -ებ-იანი მრავლობითი ფორმების ჯგუფში.

ქართული ენის ტაგსეტებში ასევე მოცემულია თანდებულიანი და ნაწილაკიანი ფორმები, თუმცა, ამ ეტაპზე მხოლოდ არსებით სახელთან.

ქართული ენის მორფოლოგიის ერთ-ერთ ყველაზე დიდ სირთულეს წარმოადგენს ზმნა, ამიტომ, მისი სრული ანოტირება გარკვეულ სიძნელეებთან იყო დაკავშირებული. ჩვენ მიერ შემუშავებული ზოგადი სტრატეგიის მიხედვით ზმნაში აღწერილია შემდეგი ჯგუფები: ფორმა: მთავარი და ბრძანებითი; პირი - პირველი, მეორე, მესამე; რიცხვი: მხოლობითი და მრავლობითი და დროები: აწმყო, მყოფადი, უწყვეტელი, ხოლმეობითი, აწმოს კავშირებითი, მყოფადის კავშირებითი, წყვეტილი, II კავშირებითი, I თურმეობითი, II თურმეობითი, III კავშირებითი.

არანაკლებ სირთულეებთანაა დაკავშირებული მასდარისა და მიმღეობის ანოტირება. უშუალოდ კონტექსტის გარეშე შეუძლებელია მასდარისა და არსებითი სახელის გარჩევა და მიმღეობისა და ზედსართავი სახელის გარჩევა. ამდენად, კორპუსში ომონიმია არ არის მოხსნილი, ამისათვის აუცილებელია კონტექსტის ნახვა და ხელით მითითება/მონიშვნა, რასაც ბუნებრივია ავტომატური მორფოლოგიური ანალიზი ვერ უზრუნველყოფს.

შედგენილი შემასმენელი ცალკე კატეგორიად არის გამოტანილი და ჯგუფებში კი მიეთითება შესაბამისი მეტყველების ნაწილები, რომლებიც შედგენილი შემასმენლის სახელად ნაწილში არიან წარმოდგენილი.

რაც შეეხება ზმნიზედას, კავშირს, შორისდებულს, თანდებულსა და ნაწილაკს, ისინი ტაგსებში მოცემულია ძირითადი კატეგორიების სახით და წარმოადგენენ ისეთ ძირითად კატეგორიებს, რომელთაც არ გააჩნიათ ატრიბუტული ჯგუფები და მათი მორფოლოგიური ანალიზი გაცილებით ადვილი იყო მეტყველების სხვა ნაწილებთან შედარებით.

მეტყველების ნაწილების გარდა ტაგსებში ცალკე კატეგორიებად წარმოდგენილია ორი დამატებითი კატეგორია: „აბრევიატურა“ და „უცნობი“. აბრევიატურაში მოცემულია ცალკეული ასოები, აბრევიატურები, პუნქტუაციის ნიშნები და ა.შ. რაც შეეხება უცნობ კატეგორიას, ამ უკანასკნელში მოცემულია უცხო სიტყვები, რომელიც ზემოთ მოცემულ არც ერთ კატეგორიას არ განეკუთვნება.

### **3.2 მორფოლოგიური ანალიზი**

როგორც უკვე აღინიშნა, ქართული ენის ვებკორპუსის მორფოლოგიური ანოტირების პლატფორმად გამოყენებულია „მულტიტექსტური მორფოსინტაქსური პარამეტრები“, ვერსია მეოთხე, რომლის ბაზაც მოიცავს 16 ენას. აღნიშნულ ლინგვისტურ ბაზას დაემატება ქართული ენის პარამეტრები, მას შემდეგ, რაც დასრულდება ამ უკანასკნელზე მუშაობა.

მორფოლოგიური ანალიზისათვის გამოყენებულია კრის ბიმანის (Biemann 2010) პროგრამა „ვებანო“ (WebAnno 2.0), რომელიც უზრუნველყოფს, როგორც ხელით, ასევე ავტომატურ მორფოლოგიურ ანოტირებას (Unsupervised Part-of-Speech Tagging). იხილეთ სურათი 10.



## სურათი 10: ქართული ენის ვებკორპუსი WebAnno-ში

The screenshot shows the 'Projects Settings' page in WebAnno. The 'General' tab is selected, displaying the following information:

- Name:** KaWaC
- Description:** Web Corpus for Georgian
- Project Types:**  annotation,  automation,  correction

The left sidebar lists various projects, with 'KaWaC' highlighted. Below the list are buttons for 'Create project' and 'Import Project' (with a 'Choose Files' button and 'No file chosen' text).

აღნიშნულ პროგრამაში გაერთიანდა ქართული ენისათვის შექმნილი ტაგსეტები. იხილეთ სურათი 11.

## სურათი 11: 8 ტაგსეტები WebAnno-ში

The screenshot shows the 'Tagsets' tab in the 'Projects Settings' page. The 'Tagsets' list includes: BART, MyPOS, NER\_WebAnno, Tiger, TuebaDZ. The 'Tag details' section for 'Vi3pi' is shown below:

- Name:** Vi3pi
- Description:** Verb form=indicative Person=third Number=plural Tense=imperfect

The left sidebar shows the project list with 'KaWaC' selected. The 'Import tagsets' section is also visible, showing a 'Format' dropdown set to 'JSON' and a 'Files' section with 'Choose Files' and 'No file chosen' text.

ბუნებრივი ენის დამუშავებაში (NLP – Natural Language Processing) ფართოდ გამოიყენება გარკვეულ წესებზე დაფუძნებული და სტატისტიკური მეტყველების ნაწილების მარკერების ამომცნობი პროგრამა – ტაგერი (POS Taggers), რომლისთვისაც უნდა მომზადდეს გარკვეული რაოდენობის ხელით ანოტირებული მასალა, რის შემდეგაც აღნიშნული პროგრამა იძლევა ავტომატური ანოტირების შესაძლებლობას. ქართული ენის ვებკორპუსში ხელით ანოტირებულია ყველაზე ხშირად ხმარებული 5000 სიტყვა (იხილეთ სურათი 12), რომელიც სპეციალურ ფორმატში ინტეგრირდა კრის ბიმანის ანოტირების პროგრამაში.

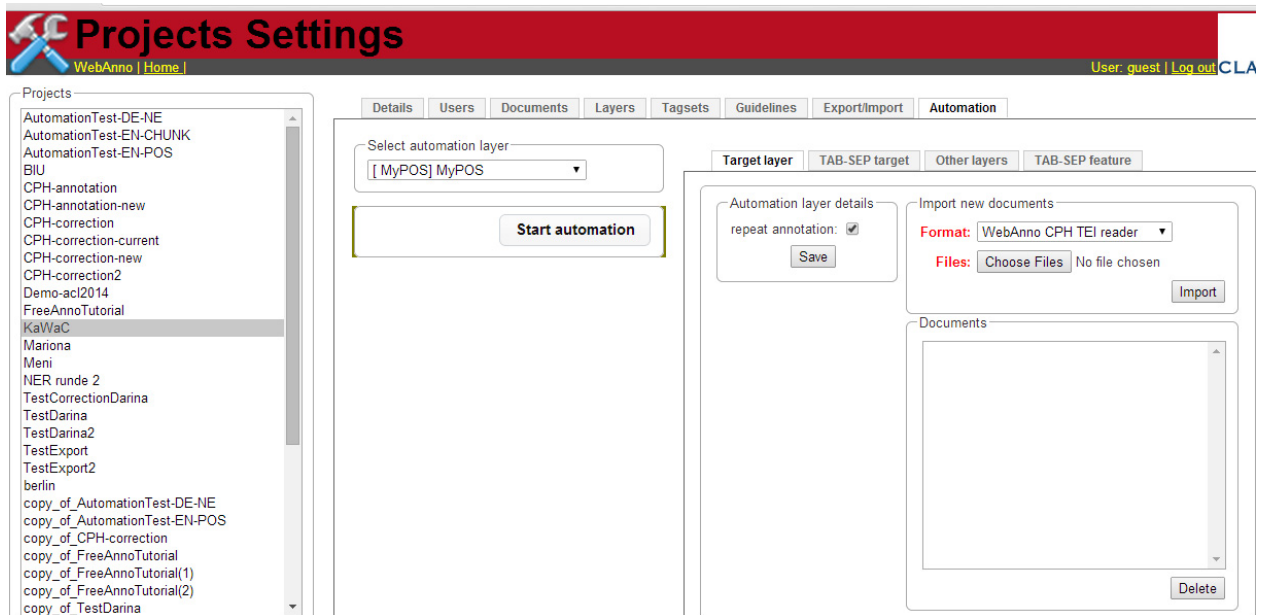
სურათი 12: ზმნის ხელით ანოტირების პროცესი

A	B	C
1	<b>MSD Features</b>	<b>Examples</b>
2	V1sp Verb form=indicative Person=first Number=singular Tense=present	ვარ, მიზნა, ვიცი, მაქვს, ვფიქრობ, მიყვარს, მგონია, მომწონს, მახსოვს, მესმის, ვამბო
3	V1pp Verb form=indicative Person=first Number=plural Tense=present	გვაცხვს, ვართ, შეგახსენებთ, ვთქვით, ვიცით, გვინდა, გთხოვთ, გირჩევთ, გვყავს, ვაწარ
4	V2sp Verb form=indicative Person=second Number=singular Tense=present	ხარ, იცი, გაქვს, ხედავ, წერ, ფიქრობ, გიყვარს, გინდა, შეგიძლია, გირჩევნია, გიჯობს, :
5	V2pp Verb form=indicative Person=second Number=plural Tense=present	ხართ, იცით, გინდათ, ფიქრობთ, ხედავთ, გვეწერთ, გგონიათ, გახსოვთ, შეგიძლიათ, გ
6	V3sp Verb form=indicative Person=third Number=singular Tense=present	არის, აქვს, არსებობს, ხდება, ამბობს, აცხადებს, წიშნავს, იცის, ჩანს, მიმდინარეობს, წ
7	V3pp Verb form=indicative Person=third Number=plural Tense=present	არიან, აქვთ, ამბობენ, აცხადებენ, იციან, ცდილობენ, უხდათ, იყენებენ, ითხოვენ, არს
8	V1sf Verb form=indicative Person=first Number=singular Tense=future	ვიტყვი, ვიქნები, დაწერ, წავალ, ვიფიქრებ, მომეწონება, შევძლებ, ეიმუშავებ, დავინა
9	V1pf Verb form=indicative Person=first Number=plural Tense=future	გვეტყვება, ვიქნებით, მივიღებთ, ვიტყვით, ვიქნებით, დავწერთ, წავალთ, ვიფიქრებთ, ი
10	V2sf Verb form=indicative Person=second Number=singular Tense=future	იტყვი, იქნები, დაწერ, წახვალ, იფიქრებ, მოგწონება, შეძლებ, იმუშავებ, დანახავ, გა
11	V2pf Verb form=indicative Person=second Number=plural Tense=future	ნახავთ, იტყვით, იქნებით, დაწერთ, მიიღებთ, წახვალთ, მოგწონებათ, შეძლებთ, იმ
12	V3sf Verb form=indicative Person=third Number=singular Tense=future	დაიწერება, შეუხება, გაიმართება, მიიღებს, მოხდება, გახდება, ექნება, დაიწყება, გაგო
13	V3pf Verb form=indicative Person=third Number=plural Tense=future	იქნებიან, ექნებათ, იტყვიან, შემლბუზნ, დაიწყებენ, მოუწყვეთ, წავლენ, დაწერენ, მიიღ
14	V1si Verb form=indicative Person=first Number=singular Tense=imperfect	ვიყავი, მქონდა, მიზნოდა, ვიცოდი, მეგონა, ვფიქრობდი, მიყვარდა, მყავდა, ვხედე
15	V1pi Verb form=indicative Person=first Number=plural Tense=imperfect	ვიყავით, გვექონდა, გვინდოდა, ვიცოდი, გვეგონა, ვფიქრობდით, გვიყვარდა, გვყავდ
16	V2si Verb form=indicative Person=second Number=singular Tense=imperfect	იყავი, გქონდა, გინდოდა, იცოდი, მეგონა, ფიქრობდი, გიყვარდა, გყავდა, ხედილობდი,
17	V2pi Verb form=indicative Person=second Number=plural Tense=imperfect	იყავით, გქონდათ, გინდოდათ, იცოდით, მეგონათ, ფიქრობდით, გიყვარდათ, ხედი
18	V3si Verb form=indicative Person=third Number=singular Tense=imperfect	ჰქონდა, ხედილობდა, იმყოფებოდა, მუშაობდა, იცოდა, არსებობდა, უხდოდა, ამბობდა, ც
19	V3pi Verb form=indicative Person=third Number=plural Tense=imperfect	წერდნენ, ჰქონდათ, ესწრებოდნენ, იცოდნენ, იყავით, მონაწილეობდნენ, იმყოფებო
20	V1sc Verb form=indicative Person=first Number=singular Tense=conditional	დაწერდი, შეუხებოდა, დავესწრებოდი, მეცოდინებოდა, ვიცხოვრებდი, ვიტყვოდი, მ
21	V1pc Verb form=indicative Person=first Number=plural Tense=conditional	დაწერდით, გვექნებოდა, დავესწრებოდი, გვეცოდინებოდა, ვიცხოვრებდით, ვიტყ
22	V2sc Verb form=indicative Person=second Number=singular Tense=conditional	დაწერდი, გვექნებოდა, დავესწრებოდი, გვეცოდინებოდა, იცხოვრებდი, იტყოდი, მიიღ
23	V2pc Verb form=indicative Person=second Number=plural Tense=conditional	დაწერდით, გვექნებოდათ, დავესწრებოდი, გვეცოდინებოდათ, იცხოვრებდით, იტყო
24	V3sc Verb form=indicative Person=third Number=singular Tense=conditional	იქნებოდა, გახდებოდა, მოახდინოს, დაწერდა, ექნებოდა, დაესწრებოდა, ეცოდინებო
25	V3pc Verb form=indicative Person=third Number=plural Tense=conditional	დაწერდნენ, ექნებოდათ, დავესწრებოდნენ, ეცოდინებოდათ, იცხოვრებდნენ, იტყო
26	V1sw Verb form=indicative Person=first Number=singular Tense=present subjunctive	წერდე, მქონდეს, ვესწრებოდე, ვიცოდე, ვცხოვრობდე, ვამბობდე, ვილბდე, ვიჯდე,
27	V1pw Verb form=indicative Person=first Number=plural Tense=present subjunctive	წერდეთ, გქონდეს, ვესწრებოდეთ, ვიცოდეთ, ვცხოვრობდეთ, ვამბობდეთ, ვილბ
28	V2sw Verb form=indicative Person=second Number=singular Tense=present subjunctive	წერდე, გქონდეს, ესწრებოდე, იცოდე, ცხოვრობდე, ამბობდე, ილბდე, იჯდე, ახდენ
29	V2pw Verb form=indicative Person=second Number=plural Tense=present subjunctive	წერდეთ, გქონდეთ, ესწრებოდეთ, იცოდეთ, ცხოვრობდეთ, ამბობდეთ, ილბდეთ, იჯ
30	V3sw Verb form=indicative Person=third Number=singular Tense=present subjunctive	არსებოდეს, ჰქონდეს, წაართვას, ესწრებოდეს, იცოდეს, ცხოვრობდეს, ილბდეს, იყ
31	V3pw Verb form=indicative Person=third Number=plural Tense=present subjunctive	წერდნენ, იცხოვრებდნენ, ილბდნენ, იხდნენ, ახდენდნენ,
32	V1sy Verb form=indicative Person=first Number=singular Tense=future subjunctive	დაწერდე, შეუხებოდეს, დავესწრებოდე, მეცოდინებოდეს, ვიცხოვრებდე, ვიტყოდე,
33	V1py Verb form=indicative Person=first Number=plural Tense=future subjunctive	დაწერდით, გვექნებოდეს, დავესწრებოდით, გვეცოდინებოდეს, ვიცხოვრებდით, ვი
34	V2sy Verb form=indicative Person=second Number=singular Tense=future subjunctive	დაწერდე, გვექნებოდეს, დავესწრებოდე, გვეცოდინებოდეს, იცხოვრებდე, იტყოდე, მი
35	V2py Verb form=indicative Person=second Number=plural Tense=future subjunctive	დაწერდეთ, გვექნებოდეთ, დავესწრებოდით, გვეცოდინებოდეთ, იცხოვრებდით, იტყო
36	V3sy Verb form=indicative Person=third Number=singular Tense=future subjunctive	დაწერდეს, ექნებოდეს, დავესწრებოდეს, ეცოდინებოდეს, იცხოვრებდეს, იტყოდეს, მ

აღნიშნულ პროგრამაში ქართულ ენის მეტყველების ნაწილების მარკერებთან ერთად ინტეგრირდა ხელით ანოტირებული 5000 ყველაზე ხშირად ხმარებული სიტყვა ქართული ენის ვებკორპუსის მიხედვით, რის შემდეგაც პროგრამის საშუალებით გაკეთდა ავტომატური მორფოლოგიური ანალიზი (იხილეთ სურათი 13). უნდა აღინიშნოს, რომ მორფოლოგიური ანოტირების მეორე ეტაპი ჩატარდა უფრო ვრცელი ლექსიკონის საფუძველზე, კერძოდ, 35 000 საწვრთნელი მასალის (training data) ბაზაზე. 35,000-სიტყვიანი საწვრთნელი

მასალის შედეგად განხორციელდა ქართული ენის ვებკორპუსის ანოტირება, რის შემდეგმაც კორპუსიდან ამოვიღეთ დაახლოებით 1 მილიონი ავტომატურად ანოტირებული სიტყვა, რომელიც შემოწმდა ხელით და გამოვლინდა ის სტატისტიკური შეცდომები, რომლებიც ქართული ენის ვებკორპუსის პირველ ანოტირებულ ვერსიაშია.

სურათი 13: ავტომატური ანოტირება WebAnno-ში



როგორც ზემოთ აღვნიშნეთ, მორფოლოგიური ანოტირების პირველ საფეხურზე ქართული ენის ვებკორპუსიდან ამოვიღეთ სისშირული სია - ყველაზე ხშირად ხმარებული 5,000 ქართული სიტყვა და განხორციელდა ამ მასალის ხელით ანოტირება. მეორე ეტაპზე კი დაახლოებით ერთი მილიონი ანოტირებული წინადადებების ბაზაზე მომზადდა საწვრთნელი მასალა კორპუსის ავტომატური მორფოლოგიური ანოტირებისათვის და მიღებული საწვრთნელი მასალით კი განხორციელდა მთლიანი ვებკორპუსის მორფოლოგიური ანოტირება.

ამგვარად, ლიდის უნივერსიტეტში განხორციელებული სადოქტორო კვლევის მეორე ეტაპზე მოვახდინეთ მთლიანი ვებკორპუსის ავტომატური მორფოლოგიური ანოტირება და შედეგად მივიღეთ ანოტირებული ვებკორპუსის პირველი ვერსია.

მორფოლოგიური ანოტირების შეცდომების გამოსავლენად გამოვიკვლიეთ ავტომატურად ანოტირებული მასალა (დაახლოებით 40 000 ავტომატურად

ანოტირებული წინადადება, იხილეთ სურათი 14) და გამოვლენილი შეცდომების ტიპურ ანალიზებს ქვემოთ წარმოგიდგინებ.

### სურათი 14: ავტომატური ანოტირების შედეგები

	A	B	C	
94	113	2 Ncsg (1)	პოლიტიკის(Ncsg), პროგრამების	
95	114	2 Npsn (2)	განო(Npsn), ბარო(Npsn)	
96	115	2 Vi2pp (1), Vi2sp (1)	გინდა(Vi2sp), ვინდათ(Vi2pp)	
97	116	2 Ncsdz (2)	პრესკონფერენციაზე(Ncsdz), ბრიფინგზე(Ncsdz)	
98	117	2 An (1), Gn (1)	სწორი(An), სრულყოფილი(Gn)	
99	119	2 R (1), Psn (1)	გინ(Psn), სად(R)	
100	122	2 Ncse (2)	მთავრობამ(Ncse), პარლამენტმა(Ncse)	
101	123	3 Q (1), Vi3sc (1)	იქნება(Q), იქნებოდა(Vi3sc), იქ-გა	
102	124	4 Pnsgc (1), Pse (1), Psn (1), Ncsn (1)	ზოგიერთი(Psn), ბატონი(Ncsn), რომელიღაც(Pnsgc), ზოგიერთმა(Pse)	
103	126	16 Vi3sp (2), Vi1sp (2)	ვამბო(Vi1sp), შუადა(Vi1sp), აღიარებს(Vi3sp), სჯერა(Vi3sp), დასტურდება, ადასტურებენ, აღიარებენ, სჯერათ, მწამს, ვწამობ, სჯეროდა, შუ	
104	127	2 Vi3sa (2), Vi2so (1)	თქვა(Vi2so/Vi3sa), ზრანა(Vi3sa)	
105	129	2 Ncsd (2)	შვილს(Ncsd), ემარს(Ncsd)	
106	130	11 Npsd (11)	მისი(Npsd), მარტს(Npsd), თებერვალს(Npsd), ოქტომბერს(Npsd), დეკემბერს(Npsd), ნოემბერს(Npsd), იანვარს(Npsd), სექტემბერს(Npsd), პარი	
107	131	6 Psgg (1), Psdz (1), Ppdh (1), Psi (1)	მასზე(Psdz), მისგან(Psgg), მათში(Ppdh), სამომავლოდ, მისდამი, მისით(Psi)	
108	132	2 Ppn (1)	ჩვენ(Ppn), ჩვენ	
109	133	3 Ncsn (1)	თარიღი(Ncsn), პოსტები, შიტები	
110	134	2 Wsi (1)	მიხედითაც(Wsi), თანახმადაც	
111	135	2 An (1)	საზოგადოებრივი(An), ქურდული	
112	137	4 Ncsg (1)	კოდექსის(Ncsg), დაუკანონდათ, #, არაუმეტეს	
113	138	2 Gn (1)	დამოუკიდებელი(Gn), ასახული	
114	139	4 Ncsg (2), An (1)	საბოლოო(An), მედიის(Ncsg), დასვლეთის(Ncsg), ყველანაირად	
115	140	3 Zg (1), Vi3sp (1), Vi3sa (1)	განაცხადა(Vi3sa), წერს(Vi3sp), ნათქვამია(Zg)	
116	141	3 Ncsn (2)	გოგონა(Ncsn), მამაკაცი(Ncsn), ბიჭუნა	
117	143	2 Ncsn (1)	დამოუკიდებლობა(Ncsn), დამოუკიდებლობას	
118	145	21 Vi2pf (2), Vi3sp (2), Vi3pp (1), Vi1pf (1)	ხედება(Vi3sp), ხედავს(Vi3sp), ხედვიან(Vi3pp), ნახავთ(Vi2pf), ხედავენ, ხედებოდა, ვხედებით, ხედავდა, ვხედავდი, გაიგებს, იგებს, ნახავ, ვ	
119	146	2 R (1)	ახლაც(R), ეხლაც	
120	147	3 Npsdt (1)	რუსეთთან(Npsdt), რუსეთისგან, მოსკოვთან	
121	148	4 Psn (2)	ესეც(Psn), ამასაც, ეგეც(Psn), ამაზეც	
122	149	2 Npsv (1)	ვლადიმერ(Npsv), ვლადიმირ	
123	150	2 Ncsn (1)	მასალა(Ncsn), პარამეტრები	
124	151	2 Ncsn (2)	კითხვა(Ncsn), შეიკითხა(Ncsn)	
125	153	2 Ncsn (1)	ფაქტი(Ncsn), გარემოება	
126	154	2 Ncsn (1)	პრეზენტაციის(Ncsn), პრეზენტაციას	
127	155	2 Gn (1), Ncsn (1)	კოდი(Ncsn), დამატებული(Gn)	
128	157	3 R (1), Psn (1), Psd (1)	ვინმე(Psn), ვინმეს(Psd), სადმე(R)	
129	158	5 a. პიოსესუალური, დონორობის, Please, კოფ		
130	159	2 Ad (2)	პირად(Ad), მშობლიურ(Ad)	
131	161	3 An (1)	სამომავლო(An), დაშინავენების, სამახსოვრო	
132	163	2 Ad (1)	უმსლეს(Ad), ტახტის	
133	164	17 Zn (1)	კაცია(Zn), ხალხია, ქალია, პროექტია, ქალაქია, დაავადება, ქართველან, წიგნია, ფილმია, ფერია, ჯგუფია, სისტემა, დღესასწაულია, იარაღი	
134	165	2 Npsd (1)	ამშ-ს(Npsd), ამშ-ს	
135	166	11 Vi3sp (3), Vi3si (1), Vi2sp (1)	სურს(Vi3sp), სურდა(Vi3si), აინტერესებს(Vi3sp), ემინია(Vi3sp), ემინიათ, აინტერესებთ, აინტერესებდა, გჭირდება(Vi2sp), ავალებდებულეს, ა	
136	168	2 Ncsg (1)	პერსონალის(Ncsg), მუნიციპალიტეტების	
137	169	5 Npsg (1), Ncsn (1)	ლარის(Npsg), ტონა(Ncsn), სთ, მეტმა, მეტრ	
138	170	13 Nnnd (1), Ncnd (1)	ბაროლონის(Nnnd) ხალხის(Ncnd) მანააპლის, პოლონის, აპლის, პოლიხის, ამერიკოლის, ინდოპოლონის, საბოლოის, მშობლიის, სპორტსმენი	

ქართული ენის ვებკორპუსის მორფოლოგიური ანოტირებისას გამოვლენილი ტიპური შეცდომებია:

- ა) **ლექსიკური უზუსტობა:** შემთხვევა, როდესაც ესა თუ ის სიტყვა მოცემულია საწვრთნელ მასალაში, მაგრამ არა იმ მარკერით, რომელიც მას უნდა ჰქონდეს მოცემულ კონტექსტში. ქვემოთ წარმოგიდგინებ მაგალითში **შეიძლება** არის ზმნა, საწვრთნელ მასალაში კი იგი წარმოგიდგინებია მხოლოდ როგორც ნაწილაკი და ავტომატური ანოტირებისას მოცემული სიტყვას აქვს არასწორი მარკერი: “Q” = ნაწილაკი:

#### დარჩა [Vi3sa]/ რამე [Psn], რისი [Psn]/ ჭამაც [Wsn]/ შეიძლება [Q]?

ამ სიტყვის სწორი მარკერი მოცემულ კონტექსტში არის Vi3sp = Verb, 3rd person, Singular, Present Tense (ზმნა, მესამე პირი, მხოლოდითი რიცხვი, აწმყო დრო).

- ბ) **უცნობი სიტყვა:** ეს არის სიტყვა, რომელიც არ არის მოცემულ საწვრთნელ მასალაში და მისი მონიშვნა ხდება ავტომატურად კონტექსტიდან

გამომდინარე. ზოგიერთ შემთხვევაში, პროგრამა ერთმანეთისაგან ვერ არჩევს საზოგადო და საკუთარ არსებით სახელებს, განსაკუთრებით მაშინ, როდესაც საუბარია ისტორიულ ადგილებზე. მაგალითად:

**„აწყურის ღეთისმშობლისა და ზარზმის ღეთისმშობლის...“**

მოცემულ მაგალითში **აწყურის** და **ზარზმის** მონიშნულია, როგორც საზოგადო არსებითი სახელები (Ncsg), მაშინ როდესაც მოცემული მაგალითისათვის სწორი მარკერია Npsg (Noun/proper/singular/genitive case).

**გ) ბუნდოვანი სიტყვა:** მორფოლოგიური ანოტირებისას ერთ-ერთ დიდ სირთულეს წარმოადგენს მასდარი და მიმღეობა, ვინაიდან კონტექსტი არასაკმარისია მასდარისა და მიმღეობის სწორად მონიშვნისათვის. მაგალითად:

**„ჩვენ არ გვეგების ამის ცოდნა“**

მოცემულ კონტექსტში **ცოდნა** არის მასდარი (მარკერი: Wsn), მაშინ როდესაც იგი ზემოთ წარმოდგენილ მაგალითში შეცდომით მონიშნულია, როგორც არსებითი სახელი (Ncsn).

გარდა მასდარისა და არსებითი სახელისა, ბუნდოვანი შემთხვევები ასევე გამოვლენილია ზმნასა და მასდარს შორის, მაგალითად:

**„და [C] შექმნის [Vi3sf] მათგან [Ppg] ისეთ [Psd] ძაღვს [Ncsd]“**

ამ მაგალითში შექმნის არის ზმნა (Vi3sf), მაგრამ მონიშნულია შეცდომით როგორც მასდარი ნათესაობით ბრუნვაში (Wsg = Masdar / singular/genitive case). მოცემული სიტყვის სწორი მარკერი კი უნდა იყოს **“Vi3sf”** (Verb / 3rd person, singular/future tense): ზმნა, მესამე პირი, მხოლოდითი რიცხვი, მყოფადი დრო.

მეორე მაგალითში კი პირიქით, მასდარი მონიშნულია, როგორც ზმნა:

**„ოჯახის [Ncsg] შექმნის [Wsg] / გადაწყვეტილების [Ncsg] / მიღება [Wsn]“**

ამ კონტექსტში შექმნის არის მასდარი, რომელსაც შეესაბამება მარკერი: Wsg (= Masdar, number = Singular, Case = Genitive): მასდარი, მხოლოდითი რიცხვი, ნათესაობითი ბრუნვა.



#### 4. სიხშირული სია

სადისერტაციო ნაშრომში ქართული ენის სიხშირულ ლექსიკისთან ერთად სიხშირული პრინციპით წარმოვადგინეთ სიხშირულ სიას (frequency list), ყველაზე ხშირად ხმარებული 17,000 სიტყვა-ფორმა (არა ლემა) ქართული ენის ვებკორპუსის მიხედვით. აღნიშნული სიხშირული სია განისაზღვრა კვლევის პირველ ეტაპზე, მორფოლოგიური ანოტირებისა და ლემატიზების განხორციელებამდე. ქართული ენის სიხშირული სია საინტერესოა იმ თვალსაზრისით, რომ იძლევა ინფორმაციას ყველაზე ხშირად ხმარებული სიტყვა-ფორმის შესახებ.

ქართული ენის სიხშირული სია, ყველაზე ხშირად ხმარებული 17,000 სიტყვა-ფორმა შესაბამისი სიხშირული ინდექსის მითითებით წინამდებარე სადოქტორო ნაშრომს დართული აქვს ელექტრონული სახით. რაც შეეხება ქართული ენის სიხშირულ ლექსიკას, ყველაზე ხშირად ხმარებულ 10,000 სიტყვას (ლემა) განხილულია წინამდებარე ნაშრომის III თავში.

\*\*\*

ამგვარად, სადისერტაციო ნაშრომის მეორე თავში წარმოდგენილია კვლევის ერთ-ერთი ძირითადი ნაწილი - ქართული ენის ვებკორპუსის – KaWaC-ის შექმნა. ნაშრომში დეტალურადაა აღწერილი აღნიშნული ვებკორპუსის შექმნის თითოეული ეტაპი, როგორცაა, 300 ფუნქციური მნიშვნელობის მქონე ქართული სიტყვის შერჩევა, რომელიც კორპუსული მასალის შეგროვების პროცესში ძალიან მნიშვნელოვანია ვებგვერდებიდან ტექსტური მასალის ჩამოტვირთვამდე არაქართული გვერდების, ანიმაციებისა და სხვა ლინგვისტურად არარელევანტური ინფორმაციის გამოსარიცხად; დეტალურადაა ასევე აღწერილი ვებგვერდების ანოტირებისა და გასუფთავების ეტაპები, რომელიც ეყრდნობა „კლინივალს“ - ერთ-ერთ უახლეს მეთოდოლოგიას ვებგვერდების გასუფთავების კუთხით. განსაკუთრებით საინტერესოა ვებკორპუსის დაბალანსების საკითხის გადაწყვეტა, რომელიც ასევე ეფუძნება უახლეს მეთოდოლოგიას „ფუნქციური ჟანრული განზომილებების“ (Functional Genre Dimensions) სახით. შესაბამისად, ქართული ენის ვებკორპუსის ჟანრული კლასიფიკაციის მეთოდოლოგიურ საფუძველს წარმოადგენს დღეს ვებკორპუსების დაბალანსებისათვის ფართოდ გამოყენებული „ფუნქციური ჟანრული განზომილებები“. ამავე თავშია წარმოდგენილი სადისერტაციო ნაშრომის კიდევ ერთი ძირითადი სიახლე,

კერძოდ, ქართული ენის მეტყველების მარკერების სისტემა, რომლის პლატფორმად გამოყენებულია „მულტიტექსტური მორფოსინტაქსური პარამეტრები“ (MULTEXT-East Morphosyntactic Specifications - MTE), მეოთხე ვერსია. ნაშრომში სრული სახითაა მოცემული მეტყველების ნაწილების მარკერების, ე.წ. ტაგსების სისტემა. ქართული ენისათვის მეტყველების ნაწილების მარკერების სისტემების „მულტიტექსტური მორფოსინტაქსური პარამეტრების“ მეოთხე ვერსიის შესაბამისად მნიშვნელოვანია იმ თვალსაზრისითაც, რომ ქართული ენის ვებკორპუსი შეიძლება შევადაროთ მორფოსინტაქსური პარამეტრების ბაზაში მოცემული 16 ენის კორპუსებს. ნაშრომში ასევე აღწერილია ქართული ენის ვებკორპუსის ავტომატური მორფოლოგიური ანოტირების პროცესი, კერძოდ, საწვრთნელი მასალის მომზადება, მორფოლოგიური ანოტირება, გაანალიზებულია მორფოლოგიური ანოტირებისას გამოვლენილი შეცდომები. გარდა ამისა, წარმოდგენილია ქართული ენის სიხშირული სია (frequency list), ყველაზე ხშირად ხმარებული 17 000 სიტყვა-ფორმა (არა ლემა) ქართული ენის ვებკორპუსის მიხედვით. აღნიშნული სიხშირული სია განისაზღვრა კვლევის პირველ ეტაპზე, მორფოლოგიური ანოტირებისა და ლემატიზების განხორციელებამდე. ქართული ენის სიხშირული სია საინტერესოა იმ თვალსაზრისით, რომ იძლევა ინფორმაციას ყველაზე ხშირად ხმარებული სიტყვა-ფორმების შესახებ, რაც ეჭვგარეშეა, მნიშვნელოვანია ენის კვლევისათვის.

### **III თავი**

## **ქართული ენის სიხშირული ლექსიკის განსაზღვრა**

### **1.1 ზოგადი მიმოხილვა**

XX საუკუნიდან მოყოლებული იცვლებოდა სასწავლო ლექსიკონების შედგენის მიდგომები და მეთოდოლოგია. განსაკუთრებით აღნიშნვის ღირსია კომპიუტერული ტექნოლოგიების განვითარების ეპოქაში კორპუსზე დაფუძნებული ლექსიკოგრაფია, მრავალმილიონიანი და მილიარდიანი კორპუსების ბაზაზე შექმნილი ლექსიკონები, სიტყვანის სიხშირული პრინციპით შერჩევა და ა.შ.

წინამდებარე თავში განვიხილავთ ქართულ-ევროპული ტიპის სასწავლო ლექსიკონების სიტყვანის განსაზღვრისათვის ლიდსის უნივერსიტეტში განხორციელებულ სამუშაოს, კერძოდ, ქართული ენის ვებკორპუსის ლემატიზებას და ქართული ენის სიხშირული ლექსიკის განსაზღვრას ვებკორპუსის მიხედვით.

### **1.2 კორპუსზე დაფუძნებული სიხშირული ლექსიკონები**

წინამდებარე სადოქტორო ნაშრომი – „ქართულ-ევროპული ტიპის ახალი სასწავლო ლექსიკონების საკითხები კორპუსული მეთოდოლოგიით“ ეფუძნება როგორც სასწავლო ლექსიკოგრაფიის, ასევე კორპუსული და სიხშირული ლექსიკონების მეთოდოლოგიებსა და პრინციპებს.

კომპიუტერული ტექნოლოგიების განვითარების ეპოქაში შეიცვალა ლექსიკონის შედგენის მიდგომები და მეთოდები. ენათა კორპუსებმა ახალი შესაძლებლობები მისცა ლექსიკოგრაფიას, კომპიუტერული პროგრამების საშუალებით შესაძლებელი გახდა მრავალმილიონიანი მასალის სწრაფად და უფრო მეტი სიზუსტით დამუშავება, რამაც, ერთი მხრივ, ლექსიკოგრაფის შრომა გააადვილა, მეორე მხრივ, ახალი შესაძლებლობები გამოავლინა ლექსიკოგრაფიული კვლევისათვის.

კომპიუტერული ტექნოლოგიების განვითარების კვალდაკვალ, ელექტრონული რესურსებისა და, ზოგადად, კორპუსისა და კომპიუტერული ლინგვისტიკის განვითარებამ ახალი გამოწვევების წინაშე დააყენა ლექსიკოგრაფია. კორპუსები ერთ-ერთ უპირველეს და ძირითად წყაროს წარმოადგენს ლექსიკოგრაფიისათვის, უფრო მეტიც, შეიქმნა ახალი



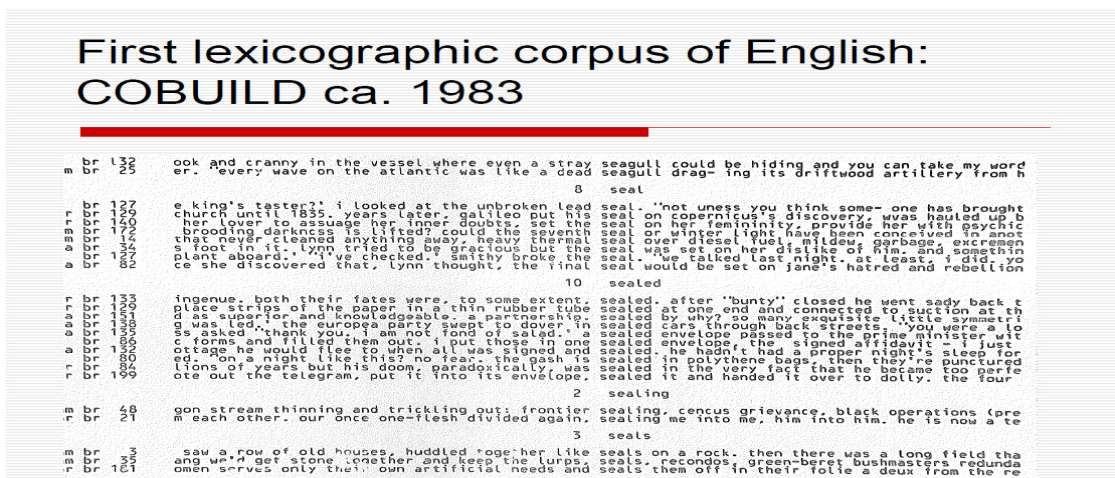
მიმართულება - კორპუსზე დაფუძნებული ლექსიკოგრაფია (corpus-based Lexicography).

იქმნება მრავალი ენის სასწავლო სისშირული ლექსიკონები მრავალმილიონიანი და მილიარდიანი კორპუსის ბაზაზე დაყრდნობით, რაც გულისხმობს ამა თუ იმ ენის ყველაზე ხშირად ხმარებული ლექსიკის განსაზღვრასა და მათ სალექსიკონო სიტყვა-სტატიებად გამოყენებას.

კორპუსზე დაფუძნებულ ერთ-ერთი პირველი ლექსიკონია კოლინზ ქობილდის ლექსიკონი, რომელიც გასული საუკუნის 90-იან წლებში გამოვიდა. აღნიშნულ ლექსიკონში სალექსიკონო ერთეულები და საილუსტრაციო მასალა კორპუსიდან ამოღებულ მასალას ეყრდნობა. კორპუსების ლექსიკოგრაფიულ წყაროდ გამოყენება დაკავშირებულია ჯონ სინკლერის სახელთან. ჯონ სინკლერმა 1980-იან წლებში ბირმინგემის უნივერსიტეტში შექმნა ინგლისური ენის კორპუსი, რომელიც იმ დროისათვის ინგლისური ენის ყველაზე ვრცელ კორპუსულ მასალას წარმოადგენდა. სინკლერის მიერ შექმნილი კორპუსი მიზნად ისახავდა ბუნებრივი, ცოცხალი ენის შესწავლასა და გაანალიზებას, რაც ქობილდის ლექსიკონის შედგენისას უნდა გაეთვალისწინებოდა. 1990-იანი წლებიდან კორპუსი გამოიყენება არა მარტო ლექსიკონებისათვის, არამედ ინგლისური ენის გრამატიკის შესწავლისათვის. ქოლინზ ქობილდის ლექსიკონის პირველად გამოცემა 1987 წელს, ხოლო ბოლო, მე-6 გამოცემა 2009 წელს.

(<http://www.collins.co.uk/category/English+Language+Teaching/COBUILD+Reference/>).

**სურათი 15: ინგლისური ენის პირველი ლექსიკოგრაფიული კორპუსი, კოლინზ ქობილდის ლექსიკონი, 1983 წ.**



რაც შეეხება თანამედროვე ლექსიკონებს, უნდა აღინიშნოს, ე.წ. „დანტე“ – Dante, რომელიც იშიფრება შემდეგნაირად: „გაანალიზებული ინგლისური ტექსტების მონაცემთა ბაზა“ (Database of Analyzed Texts of English). აღნიშნული ლექსიკონი ხელმისაწვდომია შემდეგ ონლაინ მისამართზე: <http://webdante.com/> (Atkins, Kilgarriff, Rundell 2010).

დანტეს სარედაქციო კოლეგიის შემადგენლობაში შედიოდა მსოფლიოში ყველაზე ცნობილი და ავტორიტეტული ოცი ლექსიკოგრაფი, ინგლისელი ლექსიკოგრაფის მაიკლ რანდელის საერთო რედაქტორობით. დანტე მოიცავს 42,000 სალექსიკონო სიტყვა-სტატიას, 23,000 შესიტყვებასა და 27,000-ზე მეტ იდიომასა და ფრაზას. დანტე არის მთლიანად კორპუსზე დაფუძნებული, ლექსიკონის სიტყვანი ამოღებულია 1.7-მილიარდიანი ინგლისური ენის კორპუსიდან, რომელიც ასახავს თანამედროვე ინგლისურ ენას. აღნიშნულ ლექსიკონში მოცემული საილუსტრაციო მასალაც მთლიანად კორპუსიდანაა ამოღებული.

**სურათი 16: „დანტე“ – კორპუსზე დაფუძნებული ლექსიკონი**



ასევე, არანაკლებ საინტერესოა სიტყვათშეხამებითი მოდელების პროგრამის (The Sketch Engine) მიერ შექმნილი ინოვაციური ლექსიკონი - „სკელი“ –SkELL, რაც ნიშნავს: „სიტყვათშეხამებითი მოდელების პროგრამა ენის შესწავლისათვის“ (Sketch Engine for Language Learning). „სკელი“ არის

სიტყვათშესამებითი მოდელების პროგრამაში განთავსებული მრავალმილიარდიანი კორპუსის ბაზაზე შექმნილი ლექსიკონი (Baisa, Suchomel, Kilgarriff 2015), რომელსაც იყენებენ ისეთი ავტორიტეტული გამომცემლობები, როგორებიცაა ოქსფორდის უნივერსიტეტის გამომცემლობა, კემბრიჯის უნივერსიტეტის გამომცემლობა, მაკმილანის ლექსიკონი და ა.შ. გარდა რაოდენობრივი მასალისა, შთამბეჭდავია ლექსიკონის საძიებო სისტემა, აღნიშნული ლექსიკონის ბაზაზე შესაძლებელია ცალკეული სიტყვის, სიტყვა-სიმფონიის, ე.წ. „სიტყვის-პორტრეტის“ (word-sketch), მსგავსი სიტყვების, თეზაურუსის და სხვა პარამეტრების მიხედვით ძიება. ლექსიკონი ხელმისაწვდომია შემდეგ ონლაინ მისამართზე:

<https://skell.sketchengine.co.uk/run.cgi/skell>.

### სურათი 17: „სკელი - სიტყვაშესამებითი მოდელების პროგრამა ენის შესწავლისათვის“

The screenshot shows the SKELL search engine interface. At the top, there is a search bar containing the word "corpora" and a "Search" button. To the right of the search bar are tabs for "Examples", "Word sketch", and "Similar words". Below the search bar, the results for "corpora" are displayed, showing "0.0 hits per million". A list of 24 numbered examples follows, each with the word "corpora" highlighted in red. The examples include sentences such as "Current corpora under development aim to include a hundred million words.", "America's middle class has been squeezed by corpora .", "Corpora provide a representative sample of language.", "There now exist huge volumes of text available for building corpora .", "Consequently, text corpora were limited both in size and availability.", "Such information can be found in machine readable dictionaries and corpora .", "The number of corpora lutea increases directly with age.", "This catalyzed research on methods that scale to huge corpora .", "The probabilities for these rules can be obtained from corpora .", "Rule-based linguists have traditionally made little use of corpora .", "In reference to more than one person, \"habeas corpora\".", "The rarity of such corpora reflects the effort required in the analysis.", "MAXIMIZER stimulates cell growth within the corpora cavernosa itself.", "This allows blood to flow into the corpora , producing the erection.", "The male corpora cavernosa taper off internally on reaching the spongiosum head.", "They studied corpora , \"bodies\", the medieval descendants of substances.", "The probabilities of these transitions may be more reliably obtained from available corpora .", "In order to capture as much information as possible corpora must be large.", "With such large corpora it is believed that improved language models may be created.", "One of our goals should be to develop large corpora with such data.", "The corpora collected in the first phase of this project are described below.", "The use of corpora is becoming increasingly important in the production of dictionaries.", "The blood creates pressure in the corpora cavernosa, making the penis expand.", "Natural language corpora have been available in computerised form for over thirty years."

ლექსიკოგრაფიის დღევანდელ დღეს ასახავს ახალი მიმართულება – სიხშირული ლექსიკონები (frequency dictionaries). რა განსხვავებაა სასწავლო ლექსიკონებსა და სიხშირულ ლექსიკონებს შორის? უპირველეს ყოვლისა, ამ ორ, თითქოს ერთი შეხედვით განსხვავებულ ლექსიკონებს, აერთიანებს

ლექსიკონის მკითხველი. კერძოდ, როგორც სასწავლო, ასევე სიხშირული ლექსიკონები იქმნება ამა თუ იმ ენის შესწავლით დაინტერესებული პირებისათვის და მაქსიმალურად ცდილობს ენის შემსწავლელს გაუადვილოს ენის შესწავლის პროცესი. როგორც ვიცით, სასწავლო ლექსიკოგრაფიის ჩამოყალიბების პერიოდში არ იყო ენათა კორპუსები, შესაბამისად, იგი ეყრდნობოდა როგორც გარკვეული ტექსტური მასალის სიხშირეს, ასევე მასწავლებლების პირად გამოცდილებას და ამ უკანასკნელზე დაყრდნობით განსაზღვრავდა ენის „მინიმალურ ლექსიკას“ (controlled vocabulary).

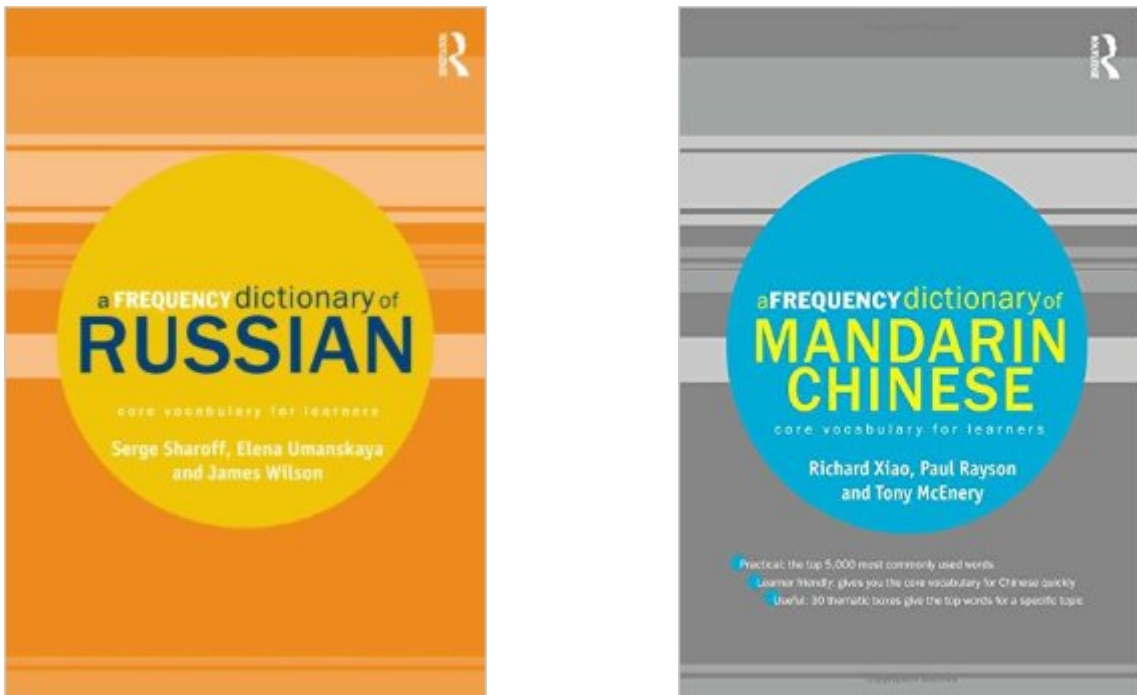
კომპიუტერებისა და ენათა კორპუსული ბაზების შექმნის კვალდაკვალ შეიცვალა აღნიშნული მიდგომები, უპირველეს ყოვლისა, შეიცვალა ტექსტური მასალის მოცულობა. მაშინ როდესაც სასწავლო ლექსიკონისათვის 1 მილიონი სიტყვის შემცველი ტექსტური მასალა განიხილებოდა საკმაოდ ვრცელ მასალად, სიხშირული ლექსიკონების შემთხვევაში საქმე გვაქვს ათეულობით მილიარდი სიტყვის შემცველ მასალაზე, რომელიც ეფუძნება ვებკორპუსებს და სიხშირული ინდექსები გამოითვლება სპეციალური კორპუსული პროგრამებით.

მაშასადამე, სასწავლო ლექსიკონებსა და სიხშირულ ლექსიკონებს შორის განსხვავება მდგომარეობს, ერთი მხრივ, ტექსტური მასალის მოცულობასა და მეორე მხრივ, ტექსტური მასალის წყაროში. სიხშირული ლექსიკონებისათვის, წყარო არის თანამედროვე მრავალმილიარდსიტყვიანი კორპუსები და რაც მთავარია, სიხშირული ინდექსები, რომლებიც სპეციალურად შექმნილი კომპიუტერული პროგრამის საშუალებით გამოითვლება.

დღეს საკმაოდ პოპულარულია სიხშირული ლექსიკონები. ამას მოწმობს ამგვარი სიხშირული ლექსიკონების სიმრავლე, რომელიც არაერთი ენისათვის არის შექმნილი და მათი რაოდენობა დღითი დღე იზრდება. სიხშირულ ლექსიკონებს შორის, შეიძლება გამოვყოთ ისეთი ლექსიკონები, როგორცაა რუსული ენის სიხშირული ლექსიკონი (Sharoff, Umanskaya, Wilson 2013), რომელსაც „რუსული ენის სიხშირული ლექსიკონი: ძირითადი სასწავლო ლექსიკა“ (A frequency dictionary of Russian: core vocabulary for learners) ეწოდება (იხილეთ სურათი 18). აღნიშნული ლექსიკონის სამუშაოებს წინ უძღოდა რუსული ენის ვებკორპუსის შექმნა და ვებკორპუსიდან სიხშირული მეთოდოლოგიით სიტყვანის დამუშავება. სიხშირული პრინციპით შეიქმნა „ჩეხური ენის სიხშირული ლექსიკონი: ძირითადი სასწავლო ლექსიკა“ (A frequency dictionary of Czech: core vocabulary for learners), (Cermák & Kren 2010,

„თანამედროვე ამერიკული ინგლისურის სიხშირული ლექსიკონი“ (A Frequency Dictionary of Contemporary American English), (Davies & Gardner 2010), „ჩინური ენის სიხშირული ლექსიკონი: ძირითადი სასწავლო ლექსიკა“ (A frequency dictionary of Mandarin Chinese: core vocabulary for learners), (Xiao, Rayson, McEnery 2009), რაც ეჭვგარეშეა, მიუთითებს იმაზე, რომ ქართულ-ევროპული ტიპის სასწავლო ლექსიკონისათვის ქართული ნაწილის სიხშირული პრინციპით დამუშავება კორპუსის ბაზაზე სიხლეს არ წარმოადგენს კორპუსზე დაფუძნებული ლექსიკოგრაფიის ისტორიაში.

**სურათი 18: რუსული და ჩინური ენის სიხშირული ლექსიკონები**



მაშასადამე, ქართული ენის სიხშირული ლექსიკის განსაზღვრა თანამედროვე და ფართოდ გამოყენებული კორპუსული მეთოდოლოგიის მიხედვით წარიმართა. უპირველეს ყოვლისა, თანამედროვე ქართული ენის ვებკორპუსი - KaWaC, შეიქმნა WaC ოჯახის კორპუსების, მაგალითად, როგორცაა ინგლისური ენის ვებკორპუსი (UkWaC), გერმანული ენის ვებკორპუსის (DeWaC), იტალიური ენის ვებკორპუსისა (ItWaC) და ა.შ. შესაბამისად. კორპუსი არის მორფოლოგიურად ანოტირებული (იხილეთ II თავი) და აღნიშნული კორპუსის ლემატიზებას და შესაბამისად, სიხშირული ლექსიკის განსაზღვრის ეტაპებს განვიხილავთ ქვემოთ, მე-2 ქვეთავში.

## 2. ქართული ენის ვებკორპუსის ლემატიზება და ქართული

### ენის სიხშირული ლექსიკის ბანსაზღვრა

ქართულ-ევროპული ტიპის სასწავლო ლექსიკონის ქართული ნაწილის განსაზღვრა კორპუსზე დაფუძნებული მეთოდოლოგიის მიხედვით წარიმართა. სიხშირული პრინციპების გათვალისწინებით ყველაზე ხშირად ხმარებული ლემების მიღებისათვის აუცილებელ სამუშაოს წარმოადგენს ლემატიზება.

ზოგადად, ლემატიზების წესები შეიძლება დაეფუძნოს ანოტირებული კორპუსიდან მიღებულ სიტყვაფორმის წყვილების შესაძლო ლემებსა და მეტყველების ნაწილების მარკერებს, ე.წ. POS (Part-of-Speech Tags) ტაგებს. ჩვენს კვლევაში გამოყენებულია CST ლემატიზების ინსტრუმენტი (CST Lemmatiser), რომელიც შემუშავებულია კოპენჰაგენის უნივერსიტეტის ენის ტექნოლოგიის ცენტრში (Københavns Universitet CST - Center for Sprogteknologi) აღნიშნული ლემატიზების ინსტრუმენტი გულისხმობს სიტყვა-ფორმების ლემატიზებას მეტყველების ნაწილების მარკერებისა და საწვრთნელი მასალის მიხედვით, რითაც ხდება კონკრეტული წესების (Sharoff & Nivre 2011) მიღება და შემდგომ მისი გამოყენება ლემატიზების პროცესში.

ქართული ენის ვებკორპუსის ლემატიზება რამდენიმე ეტაპად წარიმართა. ლემატიზების პირველ საფეხურზე კორპუსიდან ამოვიღეთ ყველაზე ხშირად ხმარებული 35,000 სიტყვა-ფორმა მორფოლოგიური ტაგებით. მეორე ეტაპზე კი კორპუსიდან ამოღებული მეტყველების ნაწილების ტაგების (POS – Part-of-Speech Tags) მქონე 35,000 სიტყვა-ფორმა ხელით დავამუშავეთ - მივუწერეთ შესაბამისი ლემები. ხელით დამუშავებული ლემატიზებული მასალის საფუძველზე მოვამზადეთ საწვრთნელი მასალა (training data) კორპუსის ავტომატური ლემატიზებისათვის. საწვრთნელი მასალის მომზადების პროცესი ილუსტრირებულია მე-19 სურათზე.



**სურათი 19: ლემატიზებისათვის საწვრთნელი მასალის მომზადების პროცესი**

1	სიტყვაფორმა	POS Tag	ლემა
2	აბსოლუტური	An	აბსოლუტური
3	აგვისტო	Npsn	აგვისტო
4	აგვისტოში	Npsdh	აგვისტო
5	აგრესიული	An	აგრესიული
6	აგრძელებ	Vi2sp	აგრძელებს
7	აგრძელებს	Vi3sp	აგრძელებს
8	ადამიანად	Ncsr	ადამიანი
9	ადამიანადაც	Ncsrc	ადამიანი
10	ადამიანადვე	Ncsrv	ადამიანი
11	ადამიანადღა	Ncsrg	ადამიანი
12	ადამიანებად	Ncpr	ადამიანი
13	ადამიანებადაც	Ncprc	ადამიანი
14	ადამიანებადვე	Ncprv	ადამიანი
15	ადამიანებადღა	Ncprg	ადამიანი
16	ადამიანებთან	Ncpdt	ადამიანი

ლემატიზების მესამე ეტაპზე ხელით დამუშავებული 35,000-სიტყვიანი საწვრთნელი მასალისა და CST ლემატიზების ინსტრუმენტის საშუალებით მოვახდინეთ მთლიანი კორპუსის ავტომატური ლემატიზება.

ლემატიზების პროცესის მეოთხე ეტაპზე ამოვიღეთ ავტომატურად ლემატიზებული ყველაზე ხშირად ხმარებული ლემები კორპუსიდან, დაახლოებით 18,000 ლემა. მიღებული ლემები დავამუშავეთ ხელით, გავფილტრეთ და გამოვრიცხეთ ისეთი შემთხვევები, როდესაც ლემები არასწორად იყო შერჩეული სიტყვა-ფორმისათვის და შედეგად წარმოვადგინეთ 10,000 ყველაზე ხშირად ხმარებული სიტყვა.

წარმოდგენილი კვლევის შედეგად მიღებულ 10 000 ყველაზე ხშირად ხმარებულ სიტყვას მითითებული აქვს სიხშირული ინდექსი და ნაჩვენებია ამა თუ იმ კონკრეტული სიტყვის სიხშირული განაწილება (frequency distribution) კორპუსში. გარდა ამისა, აღნიშნული სიტყვანი გაწყობილია ანბანთრიგზე. რაც შეეხება ლემების წარმოდგენის პრინციპებს, სახელები მოცემულია მხოლოდობითი რიცხვის სახელობითი ბრუნვის ფორმებით; ხოლო ზმნა წარმოდგენილია მესამე პირის მხოლოდობითი რიცხვის აწმყოსა და მყოფადის მწკრივებში.

სადისერტაციო ნაშრომს ელექტრონული სახით დართული აქვს ქართული ენის სიხშირული ლექსიკა – ყველაზე ხშირად ხმარებული 10 000 სიტყვის ორი ვერსია: პირველი ვერსია დალაგებულია სიხშირული ზრდის მიხედვით, ხოლო მეორე ვერსია კი ანაბანის მიხედვით და მითითებული აქვს შესაბამისი სიხშირული ინდექსი.

## 2.1 ქართულ-ევროპული სასწავლო ლექსიკონის სიხშირული სიტყვანი

სასწავლო ლექსიკოგრაფიის ჟანრული თავისებურებებიდან და კორპუსული ლექსიკოგრაფიის თანამედროვე მეთოდებიდან და მიდგომებიდან გამომდინარე, ქართულ-ევროპული ტიპის სასწავლო ლექსიკონის სიტყვანი არის კორპუსზე დაფუძნებული და სიხშირული სიზუსტით გაანგარიშებული.

წინამდებარე ნაშრომს ელექტრონულად დართული აქვს 10,000 ყველაზე ხშირად ხმარებული ლემა გამოყენების სიხშირული ინდექსის მითითებით.

ქვემოთ მე-2 ცხრილში წარმოდგენილია თანამედროვე ქართული ენის ვებკორპუსიდან ამოღებული ყველაზე ხშირად ხმარებული 100 ლემა:

**ცხრილი 2: ყველაზე ხშირად ხმარებული 100 სიტყვა:**

№	ლემა	სიხშირული ინდექსი
1.	და	7111607
2.	რომელიც	5706879
3.	არა	2520435
4.	რომ	2493502
5.	არის	2140828
6.	ის	1444341
7.	ესა	1370205
8.	ეს	1122143
9.	უნდა	993106
10.	ოუ	844677
11.	იგი	828941
12.	წელი	815997
13.	აქვს	807681
14.	რომელი	803134



15.	კო	760292
16.	ქვეყანა	742196
17.	მისი	726083
18.	რა	713654
19.	ყველა	665431
20.	მაგრამ	632940
21.	მე	572304
22.	მათი	540353
23.	ადამიანი	519247
24.	როგორც	516063
25.	ერთი	499843
26.	ჩემი	496895
27.	რაც	471183
28.	ასე	460075
29.	შემდეგ	438533
30.	დრო	438533
31.	სხვა	397674
32.	ვერ	394483
33.	დიდი	382860
34.	ჩვენი	372094
35.	ან	362156
36.	შესახებ	359359
37.	ძალიან	357980

38.	ახალი	351152
39.	დღეს	338864
40.	მხოლოდ	325782
41.	აცხადებს	320948
42.	პირველი	319988
43.	თავისი	318370
44.	ჩვენ	317229
45.	საქმე	315953
46.	უფრო	312533
47.	თუმცა	307879
48.	შეიძლება	283857
49.	არც	281872
50.	მიერ	274041
51.	უკვე	273274
52.	შორის	257331
53.	ასეთი	254266
54.	კიდევ	253947
55.	დღე	252689
56.	ორი	251940
57.	ხელი	245454
58.	ყველაფერი	240306
59.	იცის	233595
60.	შემთხვევა	230529

61.	სელისუფლება	228520
62.	ისე	227741
63.	როცა	227414
64.	საკუთარი	226598
65.	ინფორმაცია	226255
66.	სასელმწიფო	225797
67.	ბევრი	225571
68.	თავი	219807
69.	განცხადება	219471
70.	პრეზიდენტი	218869
71.	ამბობს	217655
72.	ამისი	217602
73.	საკითხი	216152
74.	მინისტრი	214234
75.	როგორ	210890
76.	ბოლო	209597
77.	ერთად	202118
78.	თქმა	201613
79.	იტყვის	201421
80.	გამო	198036
81.	სადაც	197380
82.	მეტი	194902
83.	მაინც	188967

84.	მაშინ	187915
85.	მეორე	187235
86.	მთავარი	183680
87.	ახლა	182591
88.	ცხოვრება	182430
89.	შედგუ	179476
90.	რადგან	177427
91.	წელს	176877
92.	რამდენიმე	175025
93.	სულ	168007
94.	მერე	167844
95.	წევრი	167568
96.	ადგილი	165373
97.	კარგი	164872
98.	როდესაც	163917
99.	მთელი	162832
100.	სახლი	161247

აღნიშნული 10,000 სისწირული ლემა გაწყობილია ანბანთრიგზე (იხილეთ სურათი 20), რომელიც ასევე ელექტრონული სახით დაგურთეთ ნაშრომს. როგორც ზემოთ აღინიშნა, სახელები მოცემულია მხოლოდითი რიცხვის სახელობითი ბრუნვის ფორმებით; ზმნა მესამე პირის მხოლოდითი რიცხვის აწმეოსა და მეოფადის მწკრივებში.

სურათი 20: ქართული ენის სიხშირული ლექსიკა ანბანთრინგზე

1	Lemma	Frequency
2	აა	576
3	ააგებს	1065
4	აადულებს	290
5	აავსებს	327
6	აამოძრავებს	196
7	აამუშავებს	1228
8	აანთებს	215
9	აარიდებს	630
10	აარჩევს	232
11	აასრულებს	14637
12	აატარებს	18126
13	ააღწევს	7845
14	ააყენებს	7530
15	ააყოლებს	233
16	ააშენებს	2068
17	ააცილებს	243
18	აახდენს	12696
19	აბა	32213
20	აბანო	1054

სიხშირული ლემების ანბანის მიხედვით განლაგებით წარმოვადგინეთ ისინი სალექსიკონო ერთეულებად / სიტყვა-სტატიებად და ამით დავასრულეთ ქართულ-ევროპული სასწავლო ლექსიკონის მაკროსტრუქტურაზე მუშაობა.

## **2.2. ქართული სიხშირული ლექსიკონის სიტყვანის შედარება**

### **„ქართულ ლექსიკონთან“**

ქართული ენის ვებკორპუსის მიხედვით განსზღვრული სიხშირული ლექსიკონის რამდენიმე სიტყვა შევადარეთ გიორგი ცოცანიძის, ნანა ლოლაძის და ქეთევან დათუკიშვილის „ქართულ ლექსიკონს“. აღნიშნული ლექსიკონი გამოიცა 2014 წელს და მასში ასახული ქართული სალიტერატურო ენის აქტიური სიტყვანი, ასევე ნეოლოგიზმები. ლექსიკონი არის სასწავლო დანიშნულების და მოიცავს 25 ათასზე მეტ სალექსიკონო ერთეულს, განმარტებულია 9 ათასამდე ნაწარმოები სიტყვა და ფრაზა და მოცემულია 29 ათასამდე ზმნური ფორმა. ამრიგად, „ქართული ლექსიკონის“ მოცულობა გაცილებით ვრცელია, ვიდრე ჩვენ მიერ წარმოდგენილი ქართული ენის სიხშირული ლექსიკა.

მიუხედავად „ქართული ლექსიკონის“ მოცულობისა, ქართულ-ევროპული სასწავლო ლექსიკონის სიტყვანთან შედარების შედეგად თვალნათლივ ჩანს ლექსიკური განსხვავებები, კერძოდ, სასწავლო ლექსიკონში მოცემული სიტყვანი არ ასახავს ახალ რეალიებს, რაც ჩვენში ახალი დამკვიდრებულია. ასეთი სიტყვები არ არის მოცემული „ქართულ ლექსიკონში“, მაგალითად, როგორცაა:

*ეკომიგრანტი*

*ელ-ფოსტა*

*ეპილატია*

*ერთსქესიანი*

*ეფექტიანი*

*ექსტრადირება*

*ექსტრემისტული და ა.შ.*

საპირისპიროდ ამისა, სიხშირული პრინციპით შერჩეული ლექსიკონის სიტყვანში არ არის ასახული ისეთი ქართული სიტყვები, როგორცაა:

*ეგვტერი*

*ეპოსი*

*ეშვანი*

*ერბო*

*ერკემალი და ა.შ.*

უნდა აღინიშნოს, რომ ასეთი სიტყვები მოცემულია ქართული ენის ვებკორპუსში, თუმცა მათი სიხშირული პრინციპით შერჩევისას ვერ მოხვდა

მოცემულ 10,000 სიტყვაში, ვინაიდან ისინი ნაკლებად აქტიურად გამოიყენება დღევანდელ ყოველდღიურ მეტყველებაში.

ქართული ენის ვებკორპუსში აღნიშნული სიტყვებისათვის შემდეგი საილუსტრაციო მასალაა მოცემული:

### **ეკომიგრანტი**

სისშირული ინდექსი: 864

- *„ამ პერიოდის განმავლობაში რამდენიმე ათეული ეკომიგრანტი მივიდა სამართლებრივი კონსულტაციისათვის“;*
- *„სოფელ განთიადში მოსახლეობის უმრავლესობას აჭარბდნენ ეკომიგრანტი ოჯახები შეადგენენ“.*

### **ერთსქესიანი**

სისშირული ინდექსი: 472

- *„მათ უკვე მიაღწიეს დასავლეთ ევროპის ბევრ ქვეყანაში იმას, რომ ერთსქესიანი კავშირები სრულფასოვან „ქორწინებად“ არის აღიარებული“.*

### **ექსტრემისტული**

სისშირული ინდექსი: 610

- *„ეს არის რასისტული და ექსტრემისტული აზრებით გაჯერებული ტექსტი“;*
- *„ბოლო ხანებში განვითარებული მოვლენები მოწმობენ ტერორისტული საშიშროების ზრდას ექსტრემისტული ძალების მხრიდან“.*

### **ექსტრადირება**

სისშირული ინდექსი: 974

- *„ირაკლი ოქრუაშვილის ექსტრადირების შანსი ამ ეტაპზე არ არსებობს“;*
- *„საქართველოს ხელისუფლებამ ათიოდე წლის წინ რამდენიმე ჩეჩენი მოქალაქის ექსტრადირება მოახდინა რუსეთში“.*

### **ეპილატია**

სისშირული ინდექსი: 404

- *„ლაზერული ეპილატია უკუნაჩვენებია ზოგიერთი დაავადების დროს“;*
- *„მე თუ მკითხავ ბანჯგელიან ხელეებს ჯობია ეპილატია გაიკეთონ“.*

## **ვებგვერდი**

სისშირული ინდექსი: 11814

- „უკვე ამოქმედდა სოციალური სააგენტოს სპეციალური ვებგვერდი“;
- „თუმცა, ინტერნეტსივრცეში ჩაშვებულია მათი ვებგვერდი“.

## **ინტერნეტი**

სისშირული ინდექსი: 20100

- „მე ვდებ პირობას, რომ ყველა სოფელში იქნება ინტერნეტი“;
- „ მაგ დღეებში ინტერნეტი არ მქონდა სამწუხაროდ“.

## **ელ-ფოსტა**

სისშირული ინდექსი: 2454

- „კომპიუტერთან დაჯდა და თავისი ელ-ფოსტა გახსნა“;
- „ელ-ფოსტა შევამოწმე, რამდენიმე დოკუმენტს გადავხედე და მოვაწესრიგე“.

## **საიტი**

სისშირული ინდექსი: 27348

- „პროექტის ფარგლებში, უახლოეს მომავალში შეიქმნება საიტი“;
- „გახნდა უამრავი საინფორმაციო საიტი, სადაც წერს ყველა“.

## **ფეისბუქი**

სისშირული ინდექსი: 12434

- „საბედნიეროდ, ფეისბუქი იძლევა უსაფრთხო მოღვაწეობის საშუალებას“;
- „ჩემი პოზიციის დასაფიქსირებლად ფეისბუქი არსებობს ჩემთვის და კარგადაც ვიყენებ“;

## **ბლოგი**

სისშირული ინდექსი: 42760

- „მხოლოდ გუშინ აღმოვაჩინე თქვენი ბლოგი, სამწუხაროდ“;
- „თანამედროვე ბლოგი გულისხმობს უფრო კერძო პირთა პირად ონლაინ დღიურებს“.



## ბრენდული

სისშირული ინდექსი: 142

- „არ აქვს ღირებულება, ბრენდული სამოსის შექმნის საშუალება“;
- „ძირითადად, ბრენდული ტანსაცმლით ვიმოსებ“.

## სპამი

სისშირული ინდექსი: 458

- „ელექტრონული წერილების ტრაფიკის 80%-ზე მეტი არის სპამი“;
- „სპამი მაშინაც კი ვრცელდება, როცა თქვენ ონლაინ იმყოფებით“.

ზემოთ მოცემული საილუსტრაციო მასალა, რომელიც ამოღებულია ქართული ენის ვებკორპუსიდან, ცხადყოფს იმ ახალი რეალიების გაჩენას ჩვენს ყოველდღიურ ცხოვრებაში, რომელთა ასახვა ლექსიკონებში, განსაკუთრებით კი სასწავლო ტიპის ლექსიკონებში აუცილებლობას წარმოადგენს.

\*\*\*

ამგვარად, III თავში წარმოვადგინეთ ქართულ-ევროპული ტიპის სასწავლო ლექსიკონის ქართული ნაწილის დამუშავებისა და ლექსიკონის მაკროსტრუქტურის განსაზღვრისათვის შესრულებული სამუშაო. კერძოდ, ქართული ენის ვებკორპუსის ლემატიზება და ქართული ენის სისშირული ლექსიკონის განსაზღვრა. რაც შეეხება, უშუალოდ ქართული ენის ვებკორპუსს, როგორც წყაროს ქართულ-ევროპული ტიპის სასწავლო ლექსიკონის სიტყვა-სტატიის შესადგენად, აღნიშნულ საკითხს მიმოვიხილავთ წინამდებარე ნაშრომის IV თავში.

გარდა ამისა, მესამე თავში მიმოვიხილეთ ლექსიკოგრაფიის ახალი მიმართულება – კორპუსზე დაფუძნებული ლექსიკოგრაფია (corpus-based lexicography) და კორპუსით წარმართული ლექსიკოგრაფია (corpus-driven lexicography) და კორპუსული მეთოდებით შედგენილი ისეთი ლექსიკონები როგორცაა კოლინზ ქობილდის ლექსიკონი, რომელიც არის პირველი კორპუსზე დაფუძნებული ლექსიკონი, თანამედროვე კორპუსზე დაფუძნებული ლექსიკონებიდან განვიხილეთ „დანტე“ - „გაანალიზებული ინგლისური ტექსტების მონაცემთა ბაზა“ (Database of Analyzed Texts of English) და სიტყვათშეხამებითი მოდელების პროგრამის (The Sketch Engine) მიერ შექმნილი

ინოვაციური ლექსიკონი - „სკელი“ –SkELL, რომელიც შემდგენაირად იშიფრება „სიტყვაშეხამებითი მოდელების პროგრამა ენის შესწავლისათვის“ (Sketch Engine for Language Learning). „სკელი“ არის სიტყვათშეხამებითი მოდელების პროგრამაში განთავსებული მრავალმილიარდიანი კორპუსის ბაზაზე შექმნილი ლექსიკონი. ასევე, განვიხილეთ ახალი სიხშირული ტიპის ლექსიკონები (frequency dictionaries), „რუსული ენის სიხშირული ლექსიკონი: ძირითადი სასწავლო ლექსიკა“; „ჩეხური ენის სიხშირული ლექსიკონი: ძირითადი სასწავლო ლექსიკა“; „ჩინური ენის სიხშირული ლექსიკონი: ძირითადი სასწავლო ლექსიკა“ და ა.შ. ამავე თავში განხილულია სადისერტაციო ნაშრომის კიდევ ერთი ძირითადი სიახლე, ქართული ენის სიხშირული ლექსიკონის განსაზღვრა და ამ მიზნით შესრულებელი სამუშაოები, როგორცაა ვებკორპუსის ლემაზიტება, დეტალურადაა აღწერილი, თუ რა მეთოდოლოგიურ საფუძველს ეყრდნობა ლემატიზება და როგორ წარიმართა ქართული ენის ვებკორპუსის ლემატიზება და შედეგად ამისა ქართული ენის სიხშირული ლექსიკონის განსაზღვრა. საილუსტრაციოდ ნაშრომში წარმოდგენილია ყველაზე 100 ყველაზე ხშირად ხმარებული სიტყვა. როგორც უკვე აღვნიშნეთ, ქართული ენის სიხშირული ლექსიკა – 10, 000 ყველაზე ხშირად ხმარებული სიტყვა ელექტრონული სახით აქვს დართული სადისერტაციო ნაშრომს.

ამასთან, ქართული ენის ვებკორპუსის მიხედვით განსაზღვრული სიხშირული ლექსიკონის სიტყვანის ნაწილი შევადარეთ გიორგი ცოცანიძის, ნანა ლოლაძის და ქეთევან დათუკიშვილის „ქართულ ლექსიკონს“, შედარების საფუძველზე შეიძლება ითქვას, რომ „ქართულ ლექსიკონში“ არაა მოცემული რიგი სიტყვებისა, რომელთაც ქართული ენის სიხშირული ლექსიკა მოიცავს. სადისერტაციო ნაშრომში საილუსტრაციოდ მოტანილია ასეთი სიტყვები შესაბამისი მაგალითებით.

## IV თავი

### ქართული ენის ვებკორპუსი - KaWaC, როგორც წყარო

#### ქართულ-ევროპული ტიპის სასწავლო ლექსიკონის

#### სიტყვა-სტატიების შესადგენად

##### 1.1 ზოგადი მიმოხილვა

წინამდებარე თავში განვიხილავთ ქართული ენის ვებკორპუსს, როგორც წყაროს ქართულ-ევროპული სასწავლო ტიპის ლექსიკონის სიტყვა-სტატიების შესადგენად, კერძოდ, აღნიშნული კორპუსიდან ამოღებული საილუსტრაციო მასალა შედარებულია ქართული ენის განმარტებითი ლექსიკონისა და „ქართული ლექსიკონის“ საილუსტრაციო მასალასთან. გარდა ამისა, განვიხილავთ, ე.წ. „ენ-გრამებს“ (N-grams), რომლებიც ერთი სიტყვის მეორე სიტყვასთან უშუალო მეზობლობის სიხშირულად გაანალიზებას უკავშირდება. შესაბამისად, წარმოვადგენთ ორკომპონენტიან, სამკომპონენტიან, ოთხკომპონენტიან და ხუთკომპონენტიან „ენ-გრამებს“ სიხშირული ინდექსის მითითებით. აღნიშნული „ენ-გრამები“ ქართული ენის ვებკორპუსიდანაა ამოღებული და სიხშირულად გაანალიზებული.

##### 12. სიხშირული პრინციპით გაანალიზებული „ენ-გრამები“

###### ქართული ენის ვებკორპუსის მიხედვით

ამ ნაწილში წარმოვადგენთ სიხშირული პრინციპით გაანალიზებულ, ე.წ. „ენ-გრამებს“ ქართული ენის ვებკორპუსის მიხედვით. უპირველეს ყოვლისა, განვმარტოთ უშუალოდ ტერმინი „ენ-გრამები“, რომელიც არის ინგლისური ტერმინის “n-grams” თარგმანი. უნდა აღინიშნოს, რომ აღნიშნული ტერმინი არის არა კორპუსის ლინგვისტიკის ტერმინი, იგი კომპიუტერული ლინგვისტიკის ტერმინია და აღნიშნავს მომიჯნავე / მოსაზღვრე / მეზობელი ფონემების, სიტყვების, სიტყვათა წყვილების ერთად რეალიზების სიხშირეს.

სადოქტორო კვლევის ფარგლებში, ე.წ. ენ-გრამები განხილულია მომიჯნავე სიტყვების ერთად რეალიზების მიხედვით, კერძოდ, სიხშირულად გაანალიზებულ და წარმოვადგენთ ორკომპონენტიანი, სამკომპონენტიანი, ოთხკომპონენტიანი და ხუთკომპონენტიანი ენ-გრამები. ქვემოთ 3 ცხრილში ილუსტრირებულია თითოეული შემთხვევის რამდენიმე მაგალითი. რაც შეეხება

უშუალოდ ენ-გრამების სრულ ცხრილს, მისი მოცულობიდან გამომდინარე, იგი ელექტრონულად აქვს დართული წინამდებარე სადოქტორო ნაშრომს.

**ცხრილი 3: ორკომპონენტიანი ენ-გრამები**

<b>№</b>	<b>ორკომპონენტიანი უნიტყვებები</b>	<b>სიხშირული ინდექსი</b>
1.	რაც შეეხება	58844
2.	მსგავი სიახლეები	30300
3.	სისხლის სამართლის	24863
4.	ადამიანის უფლებათა	17738
5.	დარწმუნებული ვარ	15909
6.	სახელმწიფო უნივერსიტეტის	15857
7.	საინფორმაციო ცენტრი	15773
8.	ბოლო დროს	15033
9.	ახალი ამბები	14232
10.	საკუთარ თავს	14182
11.	საქართველოს მთავრობის	12946
12.	ხელი შეუწყოს	12155
13.	საქართველოს პრეზიდენტი	11622
14.	დიდი ხნის	11018
15.	სამაუწყებლო კომპანია	9670
16.	თავისუფლება არის	9444
17.	მნიშვნელობა აქვს	9376
18.	დემოკრატიული ფასეულობებისა	9297
19.	იმედი მაქვს	9291
20.	ფაქტობრივი ინფორმაციისა	9286

21.	იდეების გავრცელებით	9284
22.	უმრავლესობის მოსმენა	9283
23.	უახლოეს მომავალში	8975
24.	განცხადებას ავრცელებს	8823
25.	დიდი ნაწილი	8652
26.	ბოლო თვეებში	8350
27.	ვიზიტის ფარგლებში	8261
28.	სამხედრო ძალით	8248
29.	პროკურატურაში დაიბარეს	8014
30.	სენსაციური განცხადება	7981
31.	ნატოსთან ინტეგრაციის	7933
32.	იურიდიული დახმარების	7927
33.	ბიზნეს პარტნიორი	7889
34.	ფეხი აიდგა	7883
35.	ხელს უწყობს	7872
36.	ადამიანის საქციელს	7868
37.	წაართვეს ტერიტორიები	7855
38.	თვეებში მოხდა	7853
39.	გაუგზავნე მეგობარს	7847
40.	პროექტის ფარგლებში	7742
41.	წლის მანძილზე	7727
42.	რომ ვთქვათ	7716
43.	საპარლამენტო უმცირესობის	7554
44.	საკუთარ თავზე	7500
45.	დიდი მადლობა	7497

46.	გასულის წლის	7490
47.	საკუთარი თავი	7417
48.	ევროატლანტიკურ სტრუქტურებში	7373
49.	კარგი იქნება	7346
50.	ბოლო პერიოდში	6858

**ცხრილი 3.1: სამკომპონენტური ენ-გრამები**

<b>N°</b>	<b>სამკომპონენტური შემთავრებები</b>	<b>სისწორული ინდექსი</b>
1.	თბილისის სახელმწიფო უნივერსიტეტის	8388
2.	ივანე ჯავახიშვილის სახელობის	8373
3.	ამერიკის შეერთებული შტატების	8318
4.	საქართველოს ხელისუფლებამ გადაწყვიტა	7866
5.	საქართველო არ გადაუხვევს	7865
6.	ქართულ პოლიტიკაზე ზემოქმედების	7857
7.	ბოლო თვეებში მოხდა	7852
8.	ადამიანის საქციელს ჰგავს	7852
9.	ნატოსთან ინტეგრაციის კურსს	7851
10.	ქვეყანაში ბოლო თვეებში	7851
11.	გადაუხვევს ნატოსთან ინტეგრაციის	7851
12.	ევროატლანტიკურ სტრუქტურებში ინტეგრაციით	6653
13.	გრანდიოზული ფასდაკლების აქცია	5805
14.	განმეორებითი გრანდიოზული ფასდაკლების	5801
15.	სისხლის სამართლის კოდექსის	4844
16.	იხილეთ ყველა კომენტარი	4026

17.	ამერიკის შეერთებულ შტატებში	3918
18.	რომ არ იყოს	3897
19.	ადამიანის უფლებათა დაცვის	3895
20.	სხვაგვარი გამოყენების უფლებას	3719
21.	არაკომერციული მიზნების პირობით	3719
22.	ბლოგის ავტორებთან შეთანხმება	3719
23.	საქართველო და მსოფლიო	3713
24.	განცხადებას ავრცელებს	8823
25.	შინაგან საქმეთა მინისტრი	3669
26.	და ამ დროს	3659
27.	გებსაიტზე ფორუმებით სარგებლობისათვის	3484
28.	ნატო და ევროპა	3479
29.	და რა მიზანს	3469
30.	დადებით შთაბეჭდილებას ტოვებს	3447
31.	საქართველოს სისხლის სამართლის	3398
32.	შეურაცმეონ და დამამცირონ	3322
33.	სასტიკადაც რომ შეურაცმეონ	3322
34.	რომ არ ყოფილიყო	3309
35.	მინისტრის პირველი მოადგილე	3134
36.	საქართველოს მართლმადიდებელი ეკლესიის	3098
37.	და ამავე დროს	3070
38.	დიდი მნიშვნელობა აქვს	3038
39.	და არა იყო	3007
40.	მნიშვნელობა არ აქვს	2987
41.	გამოწვეულ შედეგებზე პასუხისმგებლობას	2824

42.	პასუხისმგებლობას იღებს ავტორი	2824
43.	შედგებზე პასუხისმგებლობას იღებს	2824
44.	ხელშეკრულების ხელმოწერამდე მიაღწიოთ	2712
45.	წერილისთვის ხელი მოეწერა	2488
46.	გამოჩენას და ბრძოლას	2295
47.	საზოგადოებრივ ასპარეზზე გამოჩენას	2294
48.	რომ უწევს ცხოვრება	2291
49.	თუ გინდა დარჩე	2287
50.	ეფიქრობ და ვწერ	2287

**ცხრილი 3.2: ოთხკომპონენტური ენ-გრამები**

<b>№</b>	<b>ოთხკომპონენტური შინაარსები</b>	<b>სინთეზური ინდექსი</b>
1.	ხელი შეუწეოს დემოკრატიული ფასეულობების	9284
2.	რომ საქართველო არ გადაუხვევს	7862
3.	გადაუხვევს ნატოსთან ინტეგრაციის კურსს	7851
4.	ქვეყანაში ბოლო თვეებში მოხდა	7850
5.	რომ მას ყველაფერი მოსწონს	7850
6.	ევროპულ და ევროატლანტიკურ სტრუქტურებში	6785
7.	სარგებლობისთვის აუცილებელია რეგისტრაციის გავლა	3482
8.	და რა მიზანს ემსახურებოდა	3442
9.	ჰომოსექსუალური სცენების შემცველი რომანები	3437
10.	იყო და არა იყო	2935
11.	ახალ და ძველ ხელისუფლებას	2889
12.	მიმდინარეობს წართმეული ქონების გადანაწილება	2887



13.	გამოწვეულ შედეგებზე პასუხისმგებლობას იღებს	2824
14.	კეთილი იყოს თქვენი მობრძანება	2591
15.	საზოგადოებასთან და უცხოეთთან ურთიერთობას	2463
16.	სამარცხვინოდ უწიგნური პრემიერი გვყავს	2393
17.	ღვთის უკეთესი რა იქნებოდა	2350
18.	ასპარეზზე გამოჩენას და ბრძოლას	2294
19.	ან საერთოდ რომ არსებობენ	2286
20.	თუ მაღლა ასვლა დააპირებ	2286
21.	დაცემულს გნახავენ ადამიანები შეგიცოდებენ	2286
22.	გინდა დარჩე ნორმალურ ადამიანად	2286
23.	რომ უწევთ ცხოვრება შიშით	2286
24.	სიკვდილით ემუქრება ხალხს ბუნება	2285
25.	საკუთარი თავი უნდა გავაკეთილშობილოთ	2285
26.	ვინც სიკვდილით ემუქრება ხალხს	2285
27.	გულები უნდა მოვიყვანოთ წესრიგში	2285
28.	ხალხს არ აკრთობს სიკვდილი	2285
29.	ვაპირებ ვუჩივლო საქართველოს პრემიერს	2277
30.	საქართველოს სისხლის სამართლის კოდექსის	2256
31.	დაკოპირების შემთხვევაში აუცილებელია მიეთითოს	2040
32.	საიტზე გამოქვეყნებული მასალების შინაარსზე	2040
33.	შეკითხვა არავის არ დაებადებოდა	2028
34.	რომ არ იყოს ომი	1973
35.	რომ თვალები შუბლზე ამდიოდა	1970
36.	საით მივდივართ და ოხრობისკენ	1969
37.	ტუჩი მთლიანად გაშავებული მქონდა	1968

38.	ანგარიშებზე სოლიდური თანხაც ჰქონდა	1968
39.	საჭმელიც ვჭამე და წყალიც	1968
40.	ვჭამე და წყალიც დავლიე	1968
41.	რუსეთის ხელისუფლებას უჭერს მხარს	1968
42.	გამოჩენას და ბრძოლას ვაპირებ	1962
43.	პასუხები ამომწურავი არ იყო	1962
44.	უშიშარი და სამართლიანი კაცი	1948
45.	თუ რამ სათქმელი ჰქონდა	1948
46.	თუმცა მანამადე შეგიძლიათ დატკბეთ	1808
47.	ამსახველი ფოტოები და ვიდეოები	1808
48.	განსაკუთრებული მომენტების ამსახველი ფოტოები	1808
49.	საქართველოს კანონში ცვლილებების შეტანას	1798
50.	საკუთარ თავს მიანიჭა უფლება	1720

**ცხრილი 33: ხუთკომპონენტური ენ-გრამები**

<b>№</b>	<b>ხუთკომპონენტური შუსიტყვებები</b>	<b>სისშირული ინდექსი</b>
1.	დემოკრატიული ფასეულობებისა და ინსტიტუტების განმტკიცებას	9284
2.	ფაქტობრივი ინფორმაციისა თუ იდეების გავრცელებით	9284
3.	რუსეთის ხელისუფლების ქართულ პოლიტიკაზე ზემოქმედების	7856
4.	სამხედრო ძალით რუსეთს წაართვას ტერიტორიები	7855
5.	პოლიტიკა იმ ადამიანის საქციელს ჰგავს	7852
6.	საქართველო არ გადაუხვევს ნატოსთან ინტეგრაციის	7851
7.	რაც ქვეყანაში ბოლო თვეებში მოხდა	7850

8.	ევროპულ და ევროატლანტიკურ სტრუქტურებში ინტეგრაციის	6490
9.	ჯავახიშვილის სახელობის თბილისის სახელმწიფო უნივერსიტეტის	6152
10.	ჰომესექსუალური სცენების შემცველი რომანი დაურიგეს	3437
11.	უყვარს საქართველო და ამიტომაც გამოვიცხაე	2488
12.	საზოგადოებრივ ასპარეზზე გამოჩენას და ბრძოლას	2294
13.	მაგრამ საბოლოოდ მაინც სიკეთემ გაიმარჯვა	2286
14.	თუ ძირს დაცემულს გნახავენ ადამიანები	2286
15.	თუ გინდა დარჩე ნორმალურ ადამიანად	2286
16.	ან საერთოდ რომ არსებობენ ბოროტმოქმედები	2286
17.	მიწის რომელზეც დგახარ და გეიმედება	2285
18.	საკუთარი გულები უნდა მოვიყვანოთ წესრიგში	2285
19.	ევროკავშირთან ხელშეკრულების მუხლის ამოქმედებას მიესალმებიან	2279
20.	მასალების დაკოპირების შემთხვევაში აუცილებელია მიეთითოს	2040
21.	დაკოპირების შემთხვევაში აუცილებელია მიეთითოს წყარო	2040
22.	ვინც რუსეთის ხელისუფლებას უჭერს მხარს	1968
23.	საბანკო ანგარიშებზე სოლიდური თანხაც ჰქონდა	1968
24.	რომელიც ასეთ წერილებს მოაწერს ხელს	1968
25.	საჭმელიც ვჭამე და წყალიც დავლიე	1968
26.	გაეცნო და ისე მოაწერა ხელი	1967
27.	ზურგს უკან ჭორაობა არ უყვარდა	1948
28.	უშიშარი და სამართლიანი კაცი იყო	1948
29.	ცხოვრების განსაკუთრებული მომენტების ამსახველი	1808

	ფოტოები	
30.	რაც უფრო მეტი ხმაური იქნება	1718
31.	მანქანის გასაჩერებელი ადგილი არ არის	1718
32.	და პატივს ვცემ ჩემს კულტურას	1718
33.	მარტო გეი ადღუმები არ არის	1718
34.	მაგრამ ერთი რამ არ გითქვამთ	1718
35.	ფული ექნება და კაცი ერქმევა	1717
36.	თუ თბილისში ცხოვრობს თბილისს მოუაროს	1717
37.	და მაშინ გახდება ევროპული ქვეყნები	1717
38.	ჯავახიშვილის სახელობის თბილისის სახელობის უნივერსიტეტში	1533
39.	მიმართ უარყოფითად განწყობილი არ გახლავართ	1443
40.	და თქვენთვის ეგზომ ღირებული დამოკიდებულა	1443
41.	საქართველოსა და ევროკავშირს შორის ასოცირება	1436
42.	ცუდი ეკლესიის კრიტიკაში არ არის	1178
43.	საქართველოს პარლამენტში ამ დოკუმენტის რატიფიცირება	1177
44.	ანუ ვინ და რატომ ბლოკავს	1172
45.	ძალაში შესვლამდე მისი დროებითი ამოქმედების	1160
46.	გაიართ ავტორიზაცია და გააფორმოს ხელშეკრულება	1059
47.	პრობლემური საკითხები და მანკიერი მხარეები	972
48.	სახელის სწორად წარმოთქმა არ შემეძლო	845
49.	ეროვნულ თუ უცხოურ საზოგადოებას გააცნოს	821
50.	ჯგუფელი და მეგობარი მართლაც იყო	818

ამგვარად, ამ ქვეთავში წარმოვადგინეთ ორკომპონენტიანი, სამკომპონენტიანი, ოთხკომპონენტიანი და ხუთკომპონენტიანი ენ-გრამები

ქართული ენის ვებკორპუსის მიხედვით. აღნიშნული ენ-გრამები შეიძლება გამოვიყენოთ ქართულ-ევროპული სასწავლო ლექსიკონის შესიტყვებების გაანალიზებისა და განსაზღვრისათვის და საილუსტრაციო მასალის შესარჩევად.

## **2. ქართული ენის ვებკორპუსის როგორც ქართულ-ევროპული სასწავლო ლექსიკონის წყარო**

ამ ქვეთავში განვიხილავთ და გავაანალიზებთ ქართულ-ევროპული ტიპის სასწავლო ლექსიკონის სიტყვა-სტატიისათვის საილუსტრაციო მასალის შერჩევას ქართული ენის ვებკორპუსის საფუძველზე. გარდა ამისა, აღნიშნულ მასალა შედარებულია ქართული ენის განმარტებითი ლექსიკონისა და გიორგი ცოცანიძის, ნანა ლოლაძის და ქეთევან დათუკიშვილის „ქართულ ლექსიკონის“ სიტყვა-სტატიებთან.

ქართული ენის განმარტებითი ლექსიკონი გასული საუკუნის 50-იან-60-იან წლებში შეიქმნა ქართველი ენათმეცნიერის არნოლდ ჩიქობავას საერთო რედაქტორობით. ლექსიკონის შედგენაზე და რედაქციაზე მუშაობდნენ გ. ახვლედიანი, ი. გიგინეიშვილი, ვ. თოფურია, ნ. კეცხოველი, ვ. კუპრაძე, ქ. ლომთათიძე, გ. წერეთელი და სხვა. ლექსიკონში განმარტებულია 113 000-მდე (112 949) სიტყვა, რაც, როგორც ლექსიკონის წინასიტყვაობაშია ნათქვამი არ ამოწურავს ქართული ენის ლექსიკურ სიმდიდრეს. ქართული ენის განმარტებითი ლექსიკონისათვის შეიქმნა საილუსტრაციო ფრაზებისა და წინადადებების კორპუსი, რომელიც ეფუძნება ქართული ლიტერატურას, პრესიდან და სხვა, სულ მილიონ ექვსასი ათასი ბარათს. გარდა სალიტერატურო ენისა, ლექსიკონში ვრცელი ადგილი უკავია დიალექტურ მასალას. ლექსიკონის I ტომი 1950 წელს გამოვიდა, ხოლო ბოლო, მერვე ტომი - 1964 წელს.

ქართული ენის განმარტებითი ლექსიკონის ახალი რედაქციის გამოცემა დაიწყო 2008 წელს. 2010 წელს გამოქვეყნდა ახალი რედაქციის მეორე ტომი. ლექსიკონის ახალ რედაქციას ხელმძღვანელობს საქართველოს მეცნიერებათა აკადემიის წევრ-კორესპონდენტი, ენათმეცნიერების ინსტიტუტის სამეცნიერო საბჭოს თავმჯდომარე ავთანდილ არაბული.

**2.1 ქართული ენის ვებკორპუსის მიხედვით შედგენილი სალექსიკონო სიტყვა-სტატიები და მათი შედარება ქართული ენის განმარტებით ლექსიკონთან და „ქართულ ლექსიკონთან“**

ამ ნაწილში წარმოვადგენთ ქართული ენის ვებკორპუსის მიხედვით შედგენილ სალექსიკონო სიტყვა-სტატიებს და შევადარებთ ქართული ენის განმარტებითი 8-ტომეული ლექსიკონისა და გიორგი ცოცანიძის, ნანა ლოლაძის და ქეთევან დათუკიშვილის „ქართულ ლექსიკონის“ შესაბამის სიტყვა-სტატიებს.

**1) გადამისამართება**

**ქართული ენის განმარტებითი ლექსიკონი**

**გადამისამართება (გადამისამართებისა) ხელოვნ.** მისამართის შეცვლა.

*შეიძლება ითქვას, რომ ამ გამოთავისუფლებული თანხების გადამისამართება მოხდა (პარლ.).*

აღნიშნული სიტყვა არ არის მოცემული „ქართულ ლექსიკონში“.

**ქართული ენის ვებკორპუსი**

სიხშირული ინდექსი: 223

**გადამისამართება (გადამისამართებისა) 1. სატელეფონო ზარის მეორე ხაზზე გადართვა;** ორგანიზაციებს მობილურ ქსელში ისეთი დამატებითი მომსახურებების გააქტიურების საშუალება მიეცემათ, როგორცაა ზარის გადამისამართება; ძალიან ბევრს გაუჩნდა ეჭვი, რომ მოხდა ზარების გადამისამართება; **2. ინტერნეტ ბმულზე გადასვლა;** ინფორმაციის შეგროვების შემდეგ შეიძლება მოხდეს მომხმარებლის გადამისამართება რეალურ ვებ-გვერდზე; საიტზე შესვლისას, შენი სურვილის მიუხედავად, ხდება შენი გადამისამართება ისეთ საიტებზე, რომლებზეც სხვადასხვა უხამსი რეკლამა და აფიშაა განთავსებული; **3. ინფორმაციის, დახმარების და მისთ. მოპოვების მიზნით ამა თუ იმ პირთან ან ორგანიზაციასთან დაკავშირება;** მუდმივად ხდება საჯარო ინფორმაციის გაცემის ვადების დარღვევა და ჟურნალისტების ერთი უწყებიდან მეორე უწყებაში გადამისამართება; სახელმწიფო, არასახელმწიფო თუ საერთაშორისო ორგანიზაციებში გადამისამართება; შესაბამისი კომპეტენციის უფლებამოსილ პირთან გადამისამართება; **4. ა) სხვა დაწესებულებაში გადაყვანა;** პაციენტის გადამისამართება ერთი მიმართულებიდან მეორეზე გააზრებულად უნდა მოხდეს; ტენდენცია, რომ

მაინცდამაინც ემერჯენსში მოხდეს პაციენტი და მერე განხორციელდეს გადამისამართება, არასწორია; კანონს არასრულწლოვანის გადამისამართებასთან დაკავშირებით, ახალი თავი ემატება; ბ) სხვა პარტიაში, ორგანიზაციაში გადასვლა; სულაც აღარ იქნება საჭირო ხელოვნურად რომელიმე პარტიის გადამისამართება ოპოზიციურ ფლანგზე. 5. ძალების, უნარების და მისთ. მობილიზება რაიმე მიზნისკენ; უნდა მოხერხდეს მათი ძალისხმევის გადამისამართება პოზიტიური მიზნებისკენ.

როგორც ვხედავთ, ქართული ენის განმარტებით ლექსიკონში სალექსიკონო ერთეულისათვის გადამისამართება მოცემულია მხოლოდ ერთი მნიშვნელობა, მაშინ როდესაც ქართული ენის ვებკორპუსი გაცილებით ვრცელ მასალას გვაწვდის. კორპუსის მიხედვით მოცემული სიტყვის 6 მნიშვნელობა გამოიყოფა: 1. სატელეფონო ზარის მეორე ხაზზე გადართვა, 2. ინტერნეტ ბმულზე გადასვლა; 3. ინფორმაციის, დახმარების და მისთ. მოპოვების მიზნით ამა თუ იმ პირთან ან ორგანიზაციასთან დაკავშირება; 4. ა) სხვა დაწესებულებაში გადაყვანა; ბ) სხვა პარტიაში, ორგანიზაციაში გადასვლა; 6. ძალების, უნარების და მისთ. მობილიზება რაიმეს მიმართ. რაც შეეხება „ქართულ-ლექსიკონს“, აღნიშნული სიტყვა არაა მოცემული ამ ლექსიკონში.

## 2. აბღეშული

### ქართული ენის განმარტებითი ლექსიკონი

**აბღეშული-ი** (აგღეშულისა) მიმღ. ვნებ. ნამ. 1. რაც ან ვინც ააგღეს. ქსენია ბრიგადირი... ბუდიდან აგღეშული კრუხივით კაკანებდა (ნ. დუმბ.). ერთხელ ვილაცამ დაუძახა და მიტრიალდა სულ გადაავიწყდა ჰაერში აგღეშული კაცი (გ. დონ.). 2. რაც ან ვინც აიგღეს. მაყურებლის მიერ აგღეშული მსახიობი იხტიბარს არ იტეხდა. || **უპატივცემლო, შეურაცხმყოფელი.** აგღეშული ლაპარაკი იცოდა ვაგამ (გ. ბუთხ.). გულგრილობა ერის კულტურისადმი აგღეშული დამოკიდებულების მაუწყებელია («ლიტ. საქ»). გოგოს ცოტა შეეშინდა და, შიში რომ დაეძლია ან დაემალა მაინც, აგღეშული კილოთი მიუგო (ჯ. ქარჩხ.). ფრაზ. **სასაცილოდ აგღეშული,** ვისაც დასცინიან, გამასხარავებული. აი თუნდა ავილოთ ლექტორის მიერ თითქმის სასაცილოდ აგღეშული მაგალითი (ილია). სასაცილოდ აგღეშული მიხაკა ვულს მაინც არ იტეხდა (ი. გედევ.). **აბუხად აგღეშული** ანგარიშგაუწვევლობით, არაფრად მიხნევით დამცირებული (ვინმე, რამე). ის. აბუხი. ხელმწიფის შვილი ყველას აბუხად ჰყვანდა აგღეშული (ქართ. ხალხ. ზღაპ.). საუცხოოა, როცა უწყინარი, ყველასაგან აბუხად აგღეშული ხევი თავის ძალას გიჩვენებს (კ. ხიმში). სოლომონს ათასწილ ერჩივნა აბუხად ამგდები უარი მიეღო, ვიდრე კიდევ უფრო აბუხად აგღეშულს აქ, მათხოჯში დაუსრულებლივ ეყურეუტა (რ. ჯაფ.). **მასხ'არად აგღეშული** სასაცილოდ აგღეშული, გამასხარავებული.

### „ქართული ლექსიკონი“

**აბღეშული** აგღეშულის 1. რაც ააგღეს. სახურავზე აგღეშული ბურთი. 2. არასერიოზული დაუღევარი. აგღეშული დამოკიდებულება.

### ქართული ენის ვებკორპუსი

სიხშირული ინდექსი: 544

**აბღეშული-ი** (აგღეშულისა) მიმღ. ვნებ. ნამ. 1. რაც ან ვინც აიგღეს; გამოსმაურება იყო ძალიან აგღეშული; გულსატკეპნია ისეთი აგღეშული და ცინიკური კომენტარები; სახელმწიფოსა და ქვეყნისადმი აგღეშული მოპყრობა კი არაერთხელ დაგვხვდება წინ; აგღეშული აქვთ სიწმინდეები და ღმერთის შიში



ალარა აქვთ; **2. რაც ან ვინც ააგდეს;** გამარჯვებულად ის ჩაითვლება, ვინც მალლა აგდებულ ბურთს ზედიზედ სამჯერ დაიჭერს; წაღმა-უკუღმა პოზაში გატრიალებენ და ჰაერში აგდებულს ძირს არ დაგაგდებენ; ცაში აგდებული მახვილი ურჩევნია სააბაზანოს ნახვრეტში ხრიალით ჩამავალ საპნიან წყალს; **3. ამა თუ იმ პირის, ჯგუფის და მისთ. მიერ მოგონილი, წამოწყებული რამ;** სადღაც სინდისმა შეაწუხათ, მაგათი აგდებულია ეს ამბავი და შეიძლება მაგათი მოწყობილიც; ეს იყო ჩვენს წრეში აგდებული ხუმრობა, რომელიც არ უნდა გასულიყო გარეთ; **სასაცილოდ აბღეპული,** ვისაც ან რასაც დასცინიან; ქვეყანა სასაცილოდ რომაა აგდებული, ვერ ხვდებით? ათას კაცზე მხოლოდ ერთ ლოთს თუ შეხვდებით და ისიც სასაცილოდ ჰყავთ ხოლმე აგდებული; **მასხრად აბღეპული;** სასაცილოდ აგდებული, გამასხარავებული; კანონის მოქმედების შედეგები მასხრადაა აგდებული; პირველად გრძნობდა თავს მოტყუებულად, დამცირებულად, მასხრად აგდებულად; **აბუხად აბღეპული;** ანგარიშგაუწევლობით არაფრად მიჩნეული, დამცირებული; დღეს მწერალი არის შეურაცხყოფილი და მივიწყებული, აბუხად აგდებული; ცხოვრებით დაჩაგრული და ბავშვების მიერ აბუხად აგდებული კაცი; აბუხად იყო აგდებული კონსტიტუცია და, ფაქტობრივად, ერთი ადამიანის ამბიციებზე იყო მორგებული; **სამასხაროდ / სამასხროდ აბღეპული,** სასაცილოდ აგდებული, გამასხარავებული; რასაკვირველია, ახლა სამასხროდ იქნებოდა აგდებული ამ ემაწვილებისაგან; დღევანდელ ახალგაზრდობას სამასხაროდ აქვთ აგდებული; **ბანზე აბღეპული,** უყურადღებოდ დატოვებული; აშკარად უაზრო შეკითხვაა, ამიტომაც გასაგებია რატომ იქნა ეს კითხვა ბანზე აგდებული; როგორ ვერ ვიტან ბანზე აგდებულ სიტყვებს; **აბღეპული დამოკიდებულება;** მისი ძალა ისეთი იყო, რომ ჩემმა აგდებულმა დამოკიდებულებამაც კი ვერ შეამცირა; მერამდენედ გამოამჟღავნა არაპროფესიული და აგდებული დამოკიდებულება; მოსახლეობის მიმართ ცინიკური, აგდებული დამოკიდებულება; ხელისუფლების ზერელე და აგდებული დამოკიდებულება; **აბღეპული ტონი;** აშკარად დაკრავს აგდებული ტონი! უცერემონიო და აგდებული ტონი; ვაჟი შეიშმუშნა. ქალის აგდებული ტონი არ მოეწონა; **აბღეპული პასუხი;** ძალიან კატეგორიული და აგდებული პასუხი მივიღე; ეს აღტაცება მასთან სტუმრად მისულ მეგობარს გაუზიარა, რაზეც აგდებული პასუხი მიიღო; ძალზე უხეში და აგდებული პასუხიც კი გასცეს ჟურნალისტებს; **აბღეპული ლაპარაკი;** არ მიცნობს და თავიდანვე არაფრის გულისათვის აგდებული ლაპარაკი დამიწყო; აგდებულ

ლაპარაკში დაადანაშაულა; აგდებული ლაპარაკი არის იმ ზოგიერთი წესის მიმართ, რომელიც ჩემი აზრით უბრალოდ სისულელეა; **აბღეუმი მიღვრა;** მიმტანების გაბრაზებისა და უკუპირი მიდგომის მთავარი მიზეზი ქართველი კლიენტის აგდებული მიდგომაა; **აბღეუმი საუბარი;** უნდა გიყვარდეს შემოკლებული და აგდებული საუბარი; ქსენოფობიური განცხადებები და ეროვნულ უმცირესობებზე აგდებული საუბარიც მოვისმინეთ.

სალექსიკონო სიტყვა-სტატიის **აგდებულის** მხოლოდ ორი მნიშვნელობაა გამოყოფილი ქართული ენის განმარტებით ლექსიკონში, 1) რაც ან ვინც ააგდეს და 2) რაც ან ვინც აიგდეს; რაც შეეხება ქართული ენის ვებკორპუსის მიხედვით იმავე სიტყვა-სტატიას, ქართული ენის განმარტებით ლექსიკონში მოცემული ორი მნიშვნელობის გარდა გამოიყოფა მესამე მნიშვნელობა, კერძოდ, „ამა თუ იმ პირის მიერ მოგონილი, წამოწყებული რამ“ - *სადღაც სინდისმა შეაწუხათ, მაგათი აგდებულია ეს ამბავი და შეიძლება მაგათი მოწყობილიც; ეს იყო ჩვენს წრეში აგდებული ხუმრობა, რომელიც არ უნდა გასულიყო გარეთ*; აღსანიშნავია, ისიც, რომ შეცვლილია სიტყვის მნიშვნელობის თანმიმდევრობა, კერძოდ, ქართული ენის ვებკორპუსში პირველ მნიშვნელობად არის გამოტანილი „ვინც ან რაც აიგდეს“, აღნიშნული კი ქართული ენის განმარტებით ლექსიკონში მეორე მნიშვნელობად არის გამოტანილი. ასევე, ქართული ენის განმარტებით ლექსიკონში პირველ მნიშვნელობად გამოტანილია აღნიშნული სიტყვის მნიშვნელობა „ვინც ან რაც ააგდეს“, რაც ქართული ენის ვებკორპუსის მიხედვით, მესამე, ბოლო მნიშვნელობად არის წარმოდგენილი. გარდა ამისა, ქართული ენის ვებკორპუსიდან უამრავი შესიტყვებები დასტურდება, მათ შორის, ისეთიც, რომლებიც მოცემული არაა ქართული ენის განმარტებით ლექსიკონში, მაგალითად, როგორებიცაა: სამასხაროდ / სამასხროდ აგდებული; ბანზე აგდებული; აგდებული დამოკიდებულება; აგდებული ტონი; აგდებული პასუხი; აგდებული ლაპარაკი; აგდებული მიდგომა; აგდებული საუბარი.

### 3. აბსოლუტური

#### ქართული ენის განმარტებითი ლექსიკონი

**აბსოლუტური-0 (აბსოლუტურისა) თავისთავადი; არაშეფარდებითი. ||**  
**სრული, სრულიადი, სავსებითი.** გ. მაჭავარიანი აბსოლუტური შეურიგებლობით ებრძოდა... მეცნიერების განვითარების შემაფერხებელ ყოველგვარ რეციდივს («ლიტ. საქ»). თურქმენეთის ბევრ პუნქტში ჰაერის ტემპერატურის აბსოლუტური მაქსიმუმი უახლოვდება თერმეზისა და რეპეტიკისას («ახ. კომ»). [თვითმფრინავების] მიახლოებისას მათი ხმა აბსოლუტურ ხმაურამდე გაიზარდა (ე. ფად. თარგმ.). **აბსოლუტური სმენა სმენა**, რომელიც ფაქიზად და ზუსტად აღიქვამს ყოველგვარ ტონს. **აბსოლუტური უმრავლესობა** არჩევნების დროს მონაწილეთა დიდი უმეტესობა. **აბსოლუტური მონარქია** თვითმპყრობელური, შეუზღუდავი მონარქია. **აბსოლუტური მოთხოვნილება** საზოგადოებრივი მოთხოვნების მაქსიმალური დონე მეცნიერების და ტექნიკის შესაბამისად. **აბსოლუტური ჭეშმარიტება** ამომწურავი ცოდნა ობიექტის შესახებ, საბოლოო ჭეშმარიტება. **აბსოლუტური სიდიდე** მათემატიკური ფორმულა აბსოლუტური რიცხვისა: თუ  $a > 0$ ,  $|a| = a$ , თუ  $a < 0$ ,  $|a| = -a$ . **აბსოლუტური ტემპერატურა** ტემპერატურის ათვლა ნულიდან თერმოდინამიკურ ტემპერატურულ სკალაზე. **აბსოლუტური ნული** ტემპერატურა, ათვლილი აბსოლუტური ნულიდან. **აბსოლუტური სიმაღლე** დედამიწის ზედაპირის სიმაღლის წერტილი (ალტიტუდა) ზღვის დონიდან.

#### „ქართული ლექსიკონი“

**აბსოლუტური აბსოლუტურის სრული, ყოველისმომცველი. აბსოლუტური ძალაუფლება. აბსოლუტური სიხუმე. შეცდომაა აბსოლიტური.**

#### ქართული ენის ვებკორპუსი

სიხშირული ინდექსი: 910

**აბსოლუტური-0 (აბსოლუტურისა) თავისთავადი; არაშეფარდებითი. სრული, სრულიადი, სავსებითი.** კერძო საკუთრების უფლება არ იქნება აბსოლუტური; უნდა ითქვას, რომ ეს მოსაზრებები აბსოლუტური არ არის; აქ იყო აბსოლუტური იდეოლოგიური შეუთავსებლობა; **აბსოლუტური ბიომისაწვდომობა;** ორსაათიანი ინფუზიის დროს აბსოლუტური

ბიომისაწვდომობა დაახლოებით 60%-ია; აბსოლუტური ბიომისაწვდომობა დაახლოებით 91%-ს უტოლდება; **აბსოლუტური ბიოშედწევადობა**; დეფერასიროქსის აბსოლუტური ბიოშედწევადობა ( AUC ) შეადგენს დაახლოებით 70%-ს; გლიმპერიდიისათვის დამახასიათებელია აბსოლუტური ბიოშედწევადობა; **აბსოლუტური გონი**; ჰეგელისეული აბსოლუტური გონის რეპრეზენტაცია; თვითონვე ხსნიდა, თუ რატომ არ ხდებოდა აბსოლუტური გონის ქართველში რეალიზება; **აბსოლუტური დეფიციტი**; შაქრიანი დიაბეტი ტიპი 1-ის დროს სახეზეა ინსულინის აბსოლუტური დეფიციტი; ინსულინწარმოქმნელი უჯრედების დესტრუქცია და ინსულინის აბსოლუტური დეფიციტის განვითარება; გარდა ნოსტალგიისა, ისევ და ისევ ფანტაზიის აბსოლუტური დეფიციტი; **აბსოლუტური თავისუფლება**; ამიტომაც მინდოდა, იქ ყოფილიყო აბსოლუტური თავისუფლება, კანონების გარეშე, გამოსახვის ფორმების არჩევის სრული და აბსოლუტური თავისუფლება; მიუხედავად ამისა, ადამიანების უმეტესობას აბსოლუტური თავისუფლება სწადია; **აბსოლუტური მაქსიმუმი**; ჰაერის ტემპერატურის წლიური აბსოლუტური მაქსიმუმების საშუალო 35<sup>0</sup>, ხოლო აბსოლუტური მაქსიმუმი 38<sup>0</sup>-ს აღწევს; არჩევნებზე მრგვალმა მაგიდამ ხმების აბსოლუტური მაქსიმუმი მიიღო; **აბსოლუტური მინიმუმი**; ტემპერატურის აბსოლუტური მინიმუმი შესაძლოა -19, 20<sup>0</sup>-მდე დაეცეს; ორმოცი მაინც ნებისმიერ შემთხვევაში აბსოლუტური მინიმუმია; **აბსოლუტური მნიშვნელობა**; აბსოლუტური მნიშვნელობით აღემატება წინა მაქსიმუმს, ხოლო ოსცილატორის მრუდი ამას არ ადასტურებს; **აბსოლუტური მონარქია**; ეს უნდა ყოფილიყო არა ფეოდალური აბსოლუტური მონარქია, არამედ პარლამენტური, კონსტიტუციური მონარქია; ეს არის ქვეყანა, სადაც აბსოლუტური მონარქიაა, ქაღალდზეც კი არ არის დეკლარირებული ადამიანთა უფლებებო; **აბსოლუტური ნული**; აბსოლუტური ნული (-273<sup>0</sup>C) კი სითბოსაბსოლუტური ნაკლებობაა; მის მისაღებად საჭიროა ნივთიერების ტემპერატურა დაიყვანოს თითქმის აბსოლუტური ნულის (-273.15<sup>0</sup>C) მაჩვენებლამდე; **აბსოლუტური რაოდენობა**; მონოციტოპენიის დროს მათი აბსოლუტური რაოდენობა 100/მკლ-ზე ნაკლებია; ლეიკოციტების რიცხვისა და ნეიტროფილების აბსოლუტური რაოდენობის რეგულარული კონტროლი; **აბსოლუტური სიმაღლე**; აბსოლუტური სიმაღლე ზღვის დონიდან საშუალოდ 220მ უდრის; სავენახედ ასათვისებელი ფართობის აბსოლუტური სიმაღლე ზღვის დონიდან საშუალოდ 350-დან 550 მეტრს უდრის;

**აბსოლუტური უკუჩვენება;** ოპერაციის აბსოლუტური უკუჩვენება დღეისთვის სულ რამდენიმეა; აბსოლუტური უკუჩვენება ადგილობრივი გაუტკივარებისთვის არ არსებობს; **აბსოლუტური უმრავლესობა;** საქართველოში არსებული რელიგიური უმცირესობების აბსოლუტური უმრავლესობა; პარლამენტში აბსოლუტური უმრავლესობა მჭირდება; საპარლამენტო არჩევნების აბსოლუტური უმრავლესობით მოგებას გეგმავს; **აბსოლუტური სმენა;** სიმღერა ჩემს ოჯახში ყოველთვის უღერს, რადგან თითქმის ყველას აბსოლუტური სმენა აქვს; ყველასათვის აშკარა გახდა, რომ ახალ დირიჟორს აბსოლუტური სმენა ჰქონდა; აბსოლუტური ტემპერატურა; **აბსოლუტური ტემპერატურა 300** (26,8გრადუსი) კელვინია; აირის სიმკვრივე წნევის 50%-ით გადიდებისას და აბსოლუტური ტემპერატურის 2-ჯერ შეცირებისას.

ქართული ენის განმარტებით ლექსიკონში სალექსიკონო სიტყვა-სტატია **აბსოლუტური** მოცემულია ერთი მნიშვნელობით და შესაბამისი შესიტყვებებით. მოცემული სიტყვისათვის ქართული ენის ვებკორპუსში გამოიყოფა ერთი მნიშვნელობა **სრული, სრულიადი, სავსებითი**. აღსანიშნავია, რომ ქართული ენის ვებკორპუსი მოცემული სიტყვისათვის წარმოდგენილია სხადასხვა დარგის ამსახველი შესიტყვებები, მათ შორის ისეთი, რომელიც არაა მოცემული ქართული ენის განმარტებით ლექსიკონში, მაგალითად, როგორცაა: აბსოლუტური ბიომისაწვდომობა, აბსოლუტური ბიოშელწვევალობა, აბსოლუტური გონი, აბსოლუტური დეფიციტი, აბსოლუტური თავისუფლება; აბსოლუტური მაქსიმუმი; აბსოლუტური მინიმუმი, აბსოლუტური რაოდენობა; აბსოლუტური უკუჩვენება.

#### 4. ადაპტაცია

##### ქართული ენის განმარტებითი ლექსიკონი

**ადაპტაცია (ადაპტაციისა)** [ლათ. adaptation შეგუება, შეწყობა] 1. სპეც. ცოცხალი ორგანიზმების შეგუება გარემო პირობებთან არსებობის შესანარჩუნებლად. || ორგანოების შეჩვევა გარემოსთან. თვალის (მხედველობის) ადაპტაცია თვალის შეჩვევა სინათლის გაძლიერებასა თუ კლებასთან. 2. პიროვნების შეგუება მიკრო და მაკრო სოციალურ გარემოსთან. ზოგი ადაპტაციის თანდაყოლილი მახვილი უნარის წყალობით ყოველნაირ ჯოჯოხეთს შეეგუებოდა (ჭ. ამირეჯ.). ადამიანი იწყებს ადაპტაციას თავის ახალ ხნოვანებაში (ო. პაჭკ.). 3. შეწყობა, მორგება. ტექსტის ადაპტაცია ტექსტის შემოკლება-გადაკეთება მკითხველის (მსმენელის, მაყურებლის) მომზადების დონის შესაბამისად. პოპულარული ევროპული რომანების მდარე რუსულ ადაპტაციებს კითხულობდნენ მხოლოდ დაბალი კლასები (გ. გაჩეჩ.). ბენევიცმა მოგვცა «ეგმონტის» არა სცენიური ილუსტრაცია, არამედ პიესის თანამედროვე ადაპტაცია (ნ. კაკაბ.). თანამედროვე თეატრი მივიდა შექსპირის ადაპტაციის აუცილებლობამდე (ნ. ყიას.).

##### „ქართული ლექსიკონი“

**ადაპტაცია ადაპტაციის** 1. ცოცხალი ორგანიზმის შეგუება ახალ საარსებო გარემოსთან. ადაპტაცია გაუჭირდა. 2. ლიტ. რთული ტექსტის გამარტივება, ჩვეულებრივ, სწავლების მიზნით. უცხოური ტექსტის ადაპტაცია.

##### ქართული ენის ვებკორპუსი

სიხშირული ინდექსი: 970

**ადაპტაცია (ადაპტაციისა)** 1. ამა თუ იმ გარემოსთან, პირობებთან შეგუება; ქების ღირსია, რომ ასე სწრაფად შეძლო გუნდში ადაპტაცია; მას უწევს ახალ სოციალურ გარემოსთან ადაპტაცია; 2. სპეც. ცოცხალი ორგანიზმების შეგუება გარემო პირობებთან არსებობის შესანარჩუნებლად; გარემოს ტემპერატურასთან სხეულის ტემპერატურის ადაპტაცია; შემდეგ იწყება აქაურ მელიებთან ადაპტაცია; 3. შეწყობა, მორგება; ეს არის რეგულაციების ადაპტაცია, არსებულ რეალობასთან და საერთაშორისო გამოცდილებასთან მორგება; კლიმატის ცვლილების ნეგატიური ზემოქმედების მიმართ ბუნებრივი

ეკოსისტემებისა და ეკონომიკის დარგების ადაპტაცია; რაც მთავარია, ახალ წესებთან ადაპტაცია წარმატებით დაასრულა; სასწავლო პროგრამების შემუშავება და ადაპტაცია; მდგრადი მოხმარებისა და წარმოების პრაქტიკისა და ტექნოლოგიების ადაპტაცია და დანერგვა; **ტექსტის ადაპტაცია**; ძველი ქართული ტექსტის ადაპტაცია, თანამედროვე ქართულზე; სოფოკლეს პიესის სხვადასხვა თარგმანის ტექსტის ადაპტაცია დავით გაბუნიას ეკუთვნის; ტექსტის ადაპტაცია სპეციალურად ფილმისთვის „ოსკარის“ მფლობელმა სცენარისტმა ბრაიან პელგელენდმა გააკეთა; **თვალის (მხედველობის) ადაპტაცია**; თვალის შეჩვევა სინათლის გაძლიერებასა თუ კლებასთან; თვალის ადაპტაცია , თვალს აკონტროლებს მხედველობის სიმახვილისა და მხედველობის ველის ნორმალური მახვევებლები; **ბიოლოგიური ადაპტაცია**; ხომ არსებობს ბიოლოგიური ადაპტაცია, ევოლუციის კანონები და შეგუებულობა გარემოსთან? მიზიდულობის ძალისადმი ბიოლოგიური ადაპტაცია ამით არ მთავრდება; **სოციალური ადაპტაცია**; რეაბილიტაცია და სოციალური ადაპტაცია; ყველა ადამიანს სჭირდება სოციალური ადაპტაცია და ადამიანური ურთიერთობები; შეზღუდული შესაძლებლობის სტატუსის მქონე ბავშვთა სოციალური ადაპტაცია; **გენდერული ადაპტაცია**; გენდერული ადაპტაცია მოიცავს იმ ბიო-ფსიქო-სოციალურ ფაქტორებს, რომელიც ისტორიულად ყალიბდება; გენდერული ადაპტაცია მოცემულ სინამდვილეში მამაკაცისა და ქალის არსებული ხატებისადმი, გენდერული სტერეოტიპებისადმი შეგუების პროცესი.

ქართული ენის განმარტებით ლექსიკონში, სალექსიკონო ერთეულის ადაპტაციის სამი მნიშვნელობაა წარმოდგენილი, ქართული ენის ვებკორპუსშიც სამის მნიშვნელობა გამოიყოფა, ადაპტაცია ფართოდ გამოიყენება დღეს ქართულ ენაში და ნიშნავს, ამა თუ იმ პროგრამისა თუ კანონის დანერგვას: სასწავლო პროგრამების შემუშავება და ადაპტაცია; მდგრადი მოხმარებისა და წარმოების პრაქტიკისა და ტექნოლოგიების ადაპტაცია და დანერგვა. ქართული ენის ვებკორპუსის მიხედვით შედგენილ სალექსიკონო სიტყვა-სტატიაში მოცემული სიტყვის ისეთი შესიტყვეებებია წარმოდგენილი, რომლებიც ასახული არ არის ქართული ენის განმარტებით ლექსიკონში, მაგალითად: ბიოლოგიური ადაპტაცია, სოციალური ადაპტაცია, გენდერული ადაპტაცია.

## 5. ადმინისტრირება

### ქართული ენის განმარტებითი ლექსიკონი

**ადმინისტრირება** (ადმინისტრირებისა) [ლათ. administro ვმართავ, განვაგებ] ადმინისტრაციული მმართველობის განხორციელება; ფორმალურ-ბიუროკრატიული (მხოლოდ ბრძანებებითა და განკარგულებებით) მართვა. ხელმძღვანელს უაღრესად განვითარებული უნდა ჰქონდეს ადამიანთა მიზიდვის უნარი... რათა მას შეეძლოს ადმინისტრირება («ლიტ. საქ»).

აღნიშნული სიტყვა არ არის მოცემული „ქართულ ლექსიკონში“.

### ქართული ენის ვებკორპუსი

სისშირული ინდექსი: 811

**ადმინისტრირება** (ადმინისტრირებისა) 1. მართვა; ეკონომიკის ცენტრალიზაცია და მკაცრი ადმინისტრირება; გადასახადების ადმინისტრირება განსაზღვრულია კანონით, 2. ადმინისტრაციული მმართველობის განხორციელება; აქ მთავრობასაც გაუჭირდება ადმინისტრირება; ბიზნესის ადმინისტრირება; ვურჩევდი, რაიმე კონკრეტული პროფესია, მაგალითად, ბიზნესის ადმინისტრირება, ან მარკეტინგი, ან სამართალი შეესწავლა; სტიპენდიანტებს შეუძლიათ ისწავლონ ბიზნესის ადმინისტრირება; ახალი სისტემის დანერგვით ბიზნესის ადმინისტრირება უნდა გამარტივდეს; **საბაღასახალო ადმინისტრირება**; დაახლოებით ორი-სამი თვეა, საკმაოდ გამკაცრდა საგადასახალო ადმინისტრირება; მეწარმეებისთვის კუთვნილი დღგ-ს დაბრუნების თემა და საგადასახალო ადმინისტრირება; **საბაჟო ადმინისტრირება**; მუშაობის გამოცდილება ისეთ სფეროებში აქვთ, როგორცაა საბაჟო ადმინისტრირება; შემოსავლების სამსახურმა საბაჟო მიმართულებით გააძლიერა საბაჟო ადმინისტრირება და პროცედურები.

როგორც ვხედავთ, ზემოთ მოცემული სიტყვისათვის ქართული ენის განმარტებით ლექსიკონში მოცემული მხოლოდ ერთი მნიშვნელობა, ქართული ენის ვებკორპუსის მიხედვით შედგენილ სალექსიკონო სიტყვა-სტატიაში კი მოცემული სიტყვის ორი მნიშვნელობა გამოიყოფა: 1) მართვა და 2) ადმინისტრაციული მმართველობის განხორციელება. გარდა ამისა, წარმოდგენილია ისეთი შესიტყვებები, რომლებიც ძალიან აქტიურად გამოიყენება დღეს ქართულ ენაში და მისი ასახვა აუცილებელია, მაგალითად: ბიზნესის ადმინისტრირება; საგადასახალო ადმინისტრირება, საბაჟო ადმინისტრირება. რაც შეეხება „ქართულ ლექსიკონს“, არაა მოცემული აღნიშნული სიტყვა.



## 6. აპრეზია

### ქართული ენის განმარტებითი ლექსიკონი

**აპრეზია (აპრეზიზა)** სახელი აკრეფს, აიკრიფება ზმნათა მოქმედებისა 1. **შეგროვება.** თავთავის აკრეფა მოძვილ ყანაში. გეგონებოდათ, ჩინგის ყაენის მეხარკენი მოსულანო, ისეთნაირად ბრუნავდნენ სოფელში ამ ფულის აკრეფის დროს (ე. ნინოშ.). [თათარ-მონღოლები] ხარკის აკრეფის დროს არავის ინდობდნენ (საბჭ. კავშ. ისტ.). 2. **კლავიშების მეშვეობით სპეციალურ აპარატში რაიმე ინფორმაციის შეყვანა.** ტექსტის აკრეფა. ნომრის აკრეფა.

### „ქართული ლექსიკონი“

**აპრეზია აკრეფის** (აკრეფს, აუკრეფს, აიკრიფება, ააკრეფინებს) 1. **ძირს დაყრილის აღება, მოგროვება.** ვაშლები აკრიფა. 2. **სხვადასხვა პირისთვის (ორგანიზაციისთვის) განსაზღვრული თანხის გამორთმევა შესაგროვებლად.** გადასახადების აკრეფა. მეზობლებმა ფული აკრიფეს. 3. **ციფრების, ასოებისა და სიმბოლოების შეყვანა ელექტრონულ მოწყობილობაში (კომპიუტერში ტელეფონსა და მისთ.) მეგობრის ნომერი აკრიფა. მოხსენების ტექსტი აკრიფა. სეიფის პაროლი აკრიფა. აკრეფილი. რაც აკრიფეს. აკრეფილი მარცვლები.**

### ქართული ენის ვებკორპუსი

სიხშირული ინდექსი: 211

**აპრეზია (აპრეზიზა)** სახელი აკრეფს, აიკრიფება ზმნათა მოქმედებისა 1. **შეგროვება; დასწრება ნებისმიერ მსურველს შეუძლია, უფასოა და არ მოხდება ფულის აკრეფა; ეს არ არის გადასახადების აკრეფა, არამედ ხარკის აკრეფას უფრო ჰგავს;** 2. **აღება; გარეთ რომ გავდივარ, ძირს ჩამოცვენილი ფოთლების აკრეფა და ხის ტოტებზე მათი სტეპლურით მიმაგრება მინდება; ნათიამ ძირს დაყრილი ღერების აკრეფა დაიწყო;** 3. **კლავიატურით კომპიუტერში, ტელეფონში და მისთ. რაიმე ინფორმაციის შეყვანა; თავადაც შემოძლია აკრეფა, მაგრამ სწრაფი მუშაობისთვის ხელის გაჩვევა მჭირდება; თქვენ კი არანაირი პაროლის აკრეფა არ მოგიწევთ; ტემსტის აპრეზია; ახალ ამბებში რომ დავიწყო მუშაობა და „კურიერში“ დავჯექი, ტექსტის აკრეფა, დაწერა მე მომიწია; კლავიატურაზე გამაღებით დაიწყო ტექსტის აკრეფა; ნომრის აპრეზია; ბავშვი ძიძას მივეცი და ახლა მე დავიწყო ნომრის აკრეფა; სასურველი მომსახურების მისაღებად**

აუცილებელია შიდა ნომრის აკრეფა; საკმარისი იყო ცხელი ხაზის ტელეფონის ნომრის აკრეფა; **წონის აპრეფა**; კვების სწორ რეჟიმზე გადასვლისას არ ხდება წონის აკრეფა; მშობიარობის დროს და ხანდახან შემდეგაც ქალებს წონის აკრეფა ახასიათებთ; **სიჩქარის აპრეფა**; აღმართში ცოტა უჭირს სიჩქარის აკრეფა; ახალი V6-ის ტიპის სამ ლიტრიანი ძრავი აქვს, რომელსაც სიჩქარის აკრეფა 5 წამში შეუძლია.

ქართული ენის განმარტებით ლექსიკონსა და „ქართულ ლექსიკონში“ ამ სიტყვის ორი მნიშვნელობაა წარმოდგენილი, ქართული ენის ვებკორპუსის ბაზაზე შექმნილი სიტყვა-სტატიაში კი ამ სიტყვის სამი მნიშვნელობა გამოიყოფა: 1) შეგროვება, 2) აღება და 3) კლავიატურით კომპიუტერში, ტელეფონში და მისთ. რაიმე ინფორმაციის შეყვანა. გარდა ამისა, წარმოდგენილია ამ სიტყვის ისეთი შესიტყვებები როგორცაა, **წონის აპრეფა**; **კვების სწორ რეჟიმზე გადასვლისას არ ხდება წონის აკრეფა**; **სიჩქარის აპრეფა**; აღმართში ცოტა უჭირს სიჩქარის აკრეფა. უნდა აღინიშნოს, რომ ზემოთ მოცემული შესიტყვებები ქართული ენის განმარტებით ლექსიკონში არაა ასახული.

## 7. ალერგია

### ქართული ენის განმარტებითი ლექსიკონი

**ალერგია (ალერგიისა)** [ბერძ. Allos სხვა და ergon მოქმედება] ორგანიზმის გაძლიერებული მგრძობელობა ზოგი ნივთიერებისადმი, იწვევს რაიმე დაავადებას ან ავადმყოფურ მდგომარეობას. *ფიგურ. მაგ სიტყვაზე... ალერგია მემართება (თ. ჭილ.).*

აღნიშნული სიტყვა არ არის მოცემული „ქართულ ლექსიკონში“.

### ქართული ენის ვებკორპუსი

სიხშირული ინდექსი: 1048

**ალერგია (ალერგიისა)** 1. ორგანიზმის გაძლიერებული მგრძობელობა ზოგი ნივთიერებისადმი ან საჭმლისადმი, იწვევს რაიმე დაავადებას ან ავადმყოფურ მდგომარეობას; ანალიზი კეთდება, რითაც დგინდება, კონკრეტულად რაზე აქვს ალერგია; ვვარაუდობ, რომ თქვენ დაბალ და მაღალ ტემპერატურაზე გაქვთ ალერგია; 2. *გადატ. რისამე სიძულვილი, ზიზღი; სერიოზულ და ტრადიციების მოყვარულ ქალებზე აქვს ალერგია; უბრალოდ წიგნების გადაშლაზე ალერგია მაქვს და აღარაფერს ვაკეთებ; მემლიბამენტური / ჟამლისმიერი ალერგია; ბავშვს აქვს მედიკამენტური ალერგია; მედიკამენტური ალერგია ეს ის შემთხვევაა, როცა თვითნებურად, ექიმის გარეშე ანტიალერგიული პრეპარატების მიღებაც კატეგორიულად დაუშვებელია; წამლისმიერი ალერგია იშვიათად უვითარდებათ მცირეწლოვან ბავშვებსა და ასაკოვან პირებს; კვებითი ალერგია; კვებითი ალერგია ადამიანში საკვების პირველივე მიღებისას იჩენს თავს; კვებითი ალერგია არც ძუძუთი კვების დროს არის გამორიცხული.*

ქართული ენის განმარტებით ლექსიკონში მოცემული სიტყვის მხოლოდ ერთი მნიშვნელობაა გამოტანილი, ქართული ენის ვებკორპუსის მიხედვით შედგენილ სალექსიკონო სიტყვა-სტატიაში კი ამ სიტყვის ორი მნიშვნელობა გამოვყავით: 1) ორგანიზმის გაძლიერებული მგრძობელობა ზოგი ნივთიერებისადმი ან საჭმლისადმი, იწვევს რაიმე დაავადებას ან ავადმყოფურ მდგომარეობას და 2) *გადატ. რისამე სიძულვილი, ზიზღი*; ორივე ამ მნიშვნელობისთვის მოცემულია საილუსტრაციო მასალა სიტყვა-სტატიაში. გარდა ამისა, გამოვიტანეთ ისეთი შესიტყვებები, მაგალითად, როგორცაა,

მედიკამენტური / წამლისმიერი ალერგია, კვებითი ალერგია, რაც აღსანიშნავია, არ არის ასახული ქართული ენის განმარტებით ლექსიკონში. რაც შეეხება „ქართულ ლექსიკონს“, აღნიშნული სიტყვა არაა მოცემული.

## 8. ბმული

### ქართული ენის განმარტებითი ლექსიკონი

**ბმულ-0 (ბმულისა)** მიმღ. ვნებ. ნამ. 1. რაც შეაბეს, ნაბამი. ხარი სად მოვა ბუღასთან, იხიც – გუთანში ბმულია?! (ვაჟა). 2. რაც აბია, – დაბმული. ვაქო, ვით ვაქო, რომ ენა ოცნებისაგან ბმულია?! (რ. ერისთ.). 3. სპეც. რისი ბმაც (მნიშვ. 2) განხორციელდა. დაფაზე გოგირდმეყავს ბმულები იყო გამოხაზული.

### „ქართული ლექსიკონი“

**ბმული** ბმულის რაც ბმის (4) რაც ბმის შედეგად არის მიღებული. საიტის ბმული.

### ქართული ენის ვებკორპუსი

სიხშირული ინდექსი: 4611

**ბმულ-0 (ბმულისა)** 1. სპეციალური მისამართი ვებსერვერში/ინტერნეტში; სტატიის სარედაქციო კოლეჯის მიერ გამოსაქვეყნებლად მიღების შემდეგ ავტორს ეგ ზავნება სპეციალური ბმული; სამწუხაროდ საიტს არ აქვს გადმოსაწერი ბმული; დეტალებისთვის იხილეთ სახელმწიფო შესყიდვების ერთიანი ელექტრონული სისტემის ბმული; თბილისის სახელმწიფო უნივერსიტეტის ანგარიშზე დაიდო ეს ბმული; 2. რაც შეაბეს, ნაბამი; ერთის ღვაწლის დამდები, ერთს ისტორიულ უღელში ბმული; და თითქოს შემთხვევით გაწყდა, ყელზე თეთრი მძივის ბმული; 3. სპეც. რისი ბმაც განხორციელდა; ამ ფორმას აქვს ჰექსაგონალური მესერი, სადაც ყველა ატომი კოვალენტურადაა ბმული; ამ უკანასკნელებს ითვისებენ მცენარეები და გადაყავთ ორგანულად ბმულ აზოტად.

ქართული ენის განმარტებით ლექსიკონში სალექსიკონო ერთეულისათვის ბმული გამოტანილია სამი მნიშვნელობა, 1) რაც შეაბეს, ნაბამი, 2) რაც აბია, – დაბმული და 3) რისი ბმაც განხორციელდა. ხოლო „ქართულ ლექსიკონში“ ამ

სიტყვის მხოლოდ ერთი მნიშვნელობაა მოცემული. აღნიშნულ ლექსიკონებში ასახვა ვერ პოვა ისეთმა მნიშვნელობამ, როგორცაა **სპეციალური მისამართი ვებსაიტები/ინტერნეტში; სტატიის სარედაქციო კოლეგიის მიერ გამოსაქვეყნებლად მიღების შემდეგ ავტორს ეგზავნება სპეციალური ბმული; სამწუხაროდ საიტს არ აქვს გადმოსაწერი ბმული; დეტალებისთვის იხილეთ სახელმწიფო შესყიდვების ერთიანი ელექტრონული სისტემის ბმული; თბილისის სახელმწიფო უნივერსიტეტის ანგარიშზე დაიღო ეს ბმული.** აღსანიშნავია, რომ მოცემული სიტყვა დღეს ყველაზე მეტად სწორედ ამ მნიშვნელობით გამოიყენება და ქართული ენის ვებკორპუსის მიხედვით შედგენილ სალექსიკონო სიტყვა-სტატიაში ეს მნიშვნელობა პირველ მნიშვნელობადაა გამოტანილი.

## 9. ბააჰტიურება

### ქართული ენის განმარტებითი ლექსიკონი

**ბააჰტიურება (ბააჰტიურებისა)** სახელი გაააქტიურებს, გააქტიურდება ზმნათა მოქმედებისა. ამას ძალიან დიდი მნიშვნელობა აქვს, თუნდაც მაყურებლის გააქტიურების თვალსაზრისით («ლიტ. საქ»). ბოლო წლები საოცარი შინაგანი გააქტიურების, გაჯანსაღების... წლები[ა] («ლიტ. საქ»).

### „ქართული ლექსიკონი“

**ბააჰტიურება** გააქტიურების (გაააქტიურებს, გაუაქტიურებს, გააქტიურდება) 1. უფრო მეტი აქტიურობით ამოქმედება, აქტიურად ქცევა. სტუდენტთა გააქტიურება. მუშაობა გაააქტიურეს. მტერი გააქტიურდა. 2. აპარატში სპეციალური კოდის შეყვანით რაიმეს ჩართვა. ანგარიშის გააქტიურება. მონობარათი გაააქტიურა. გააქტიურებული. რაც ან ვინც გააქტიურდა. გააქტიურებული ინფექციები.

### ქართული ენის ვებკორპუსი

სისშირული ინდექსი: 1854

**ბააჰტიურება (ბააჰტიურებისა)** სახელი გაააქტიურებს, გააქტიურდება ზმნათა მოქმედებისა; 1. აქტიური, მოქმედი გახდება; დაავადების გამომწვევი მიკრობის გააქტიურება; საბუნებისმეტყველო მეცნიერებებში ადამიანური

*რესურსის გააქტიურება; 2. ამა თუ იმ მომსახურების, სერვისის ჩართვა; ერთჯერადი კოდის შეყვანის შემდეგ „SMS განვადება“ გააქტიურდება; მობილური ტელეფონის ინტერნეტ მომსახურების გააქტიურება.*

ქართული ენის განმარტებით ლექსიკონში სალექსიკონო ერთეულისათვის **გააქტიურება** წარმოდგენილია მხოლოდ ერთი მნიშვნელობა – **აქტიური, მოქმედი გახდება**. ხოლო „ქართულ ლექსიკონში“ კი მოცემულია ორი მნიშვნელობა, მათ შორის ისეთი მნიშვნელობა, როგორცაა: **ამა თუ იმ მომსახურების, სერვისის ჩართვა; ერთჯერადი კოდის შეყვანის შემდეგ „SMS განვადება“ გააქტიურდება; მობილური ტელეფონის ინტერნეტ მომსახურების გააქტიურება**. უნდა აღინიშნოს, რომ მოცემული სიტყვის ეს მნიშვნელობა ძალიან აქტიურად გამოიყენება დღეს ქართულ ენაში და მისი ასახვა ქართული ენის ლექსიკონებში აუცილებლობას წარმოადგენს.

## 10. დაკონსერვება

### ქართული ენის განმარტებითი ლექსიკონი

**დაკონსერვება (დაკონსერვებისა)** 1. სახელი დააკონსერვებს ზმნის მოქმედებისა, – კონსერვად ქცევა (რისამე), კონსერვის გაკეთება, დამზადება. ხილის დაკონსერვება. – ხორცის დაკონსერვება. ჟურნალი... გვაცნობს საკეთების დაკონსერვების წესებს («სოფ. ცხ.»). 2. წიგნ. იგივეა, რაც კონსერვაცია. მშენებლობის დაკონსერვება.

### „ქართული ლექსიკონი“

**დაკონსერვება დაკონსერვების (დააკონსერვებს, დაუკონსერვებს, დაკონსერვდება)** 1. კონსერვის დამზადება. კიტრი დააკონსერვა. თევზი დააკონსერვეს. 2. გადატ. დაწყებული საქმის, პროცესის დროებით შეჩერება. მშენებლობის დაკონსერვება. საბადო დააკონსერვეს. დაკონსერვებული. რაც დააკონსერვეს. დაკონსერვებული ბარდა.

### ქართული ენის ვებკორპუსი

სიხშირული ინდექსი: 83

**დაკონსერვება (დაკონსერვებისა)** 1. კონსერვად ქცევა (რისამე), კონსერვის გაკეთება, დამზადება; სახლის პირობებში შეიძლება არა მარტო

ბოსტნეულის დაკონსერვება, არამედ მათი გაშრობაც; დაუშვებელია ხორცის, თევზის, მწვანილის ბინის პირობებში დაკონსერვება; **2. შეჩერება;** აი, ასეთი მეთოდებით შეძლო უმართავი პროცესების დაკონსერვება; პრობლემის დაკონსერვება არ არის პრობლემის მოგვარება; ზოგჯერ შენობის დაკონსერვება უფრო ძვირი ჯდება, ვიდრე მშენებლობის გაგრძელება; **3. დაცვა რისამე მავნე ზემოქმედებებისაგან;** ამ ისტორიული ნაშთების დაკონსერვება ახლა მაინც უნდა მოხდეს.

ქართული ენის განმარტებითი ლექსიკონსა და „ქართულ ლექსიკონში“ მოცემული სიტყვის ორი მნიშვნელობაა განხილული, ქართული ენის ვებკორპუსის მიხედვით შედგენილ სალექსიკონო სიტყვა-სტატიაში კი ამ სიტყვის სამი მნიშვნელობა გამოვყავით: **1) კონსერვად ქცევა (რისამე), კონსერვის გაკეთება, დამზადება, 2) შეჩერება და 3) დაცვა რისამე მავნე ზემოქმედებებისაგან.**

## 11. დაწინააღმართება

### ქართული ენის განმარტებითი ლექსიკონი

**დაწინააღმართება (დაწინააღმართებისა)** სახელი დააფიქსირებს, დაფიქსირდება ზმნათა მოქმედებისა.

#### „ქართული ლექსიკონი“

**დაწინააღმართება** დაფიქსირების (დააფიქსირებს, დაუფიქსირებს, დაფიქსირდება, დაუფიქსირდება) მოვლენის, ფაქტის, თვალსაზრისისა და მისთ. რაიმე სახით აღნიშვნა, აღბეჭდვა. თავისი პოზიცია დააფიქსირა.

#### ქართული ენის ვებკორპუსი

სიხშირული ინდექსი: 1673

**დაწინააღმართება (დაწინააღმართებისა)** 1. ამა თუ იმ საკითხთან დაკავშირებით საკუთარი აზრის, პოზიციის გამოხატვა; მნიშვნელოვანია თავისუფლად შეეძლოს საკუთარი აზრის დაფიქსირება; მთელი გუნდის სახელით გვთხოვს მათი პოზიციის დაფიქსირება; **2. გამოვლენა, აღნუსხვა; აღბეჭდვა; დეციმეტრული და სანტიმეტრული სიხშირის რადიოტალღების დაფიქსირება;**

მოხდება თითოეული ობიექტის ფოტოზე დაფიქსირება; როგორ წარმოგიდგენიათ ამ შეთანხმების დარღვევის დაფიქსირება; 3. მიმაგრება/ დამაგრება; გამყოფი ხაზის გასწვრივ მავთულხლართების გაგდება და დაფიქსირება; სხვადასხვა ტიპის ნემსკავის ფიქსაცია სამონტაჟო ძაფის დაფიქსირება.

ქართული ენის განმარტებით ლექსიკონსა და „ქართულ ლექსიკონში“ მოცემული სალექსიკონო ერთეულის მხოლოდ ერთი მნიშვნელობაა გამოტანილი, ასევე არაა მოცემული ვრცელი საილუსტრაციო მასალა. რაც შეეხება ქართული ენის ვებკორპუსის მიხედვით შედგენილ სალექსიკონო სიტყვა-სტატიას, გამოიყოფა სამი მნიშვნელობა; 1) **ამა თუ იმ საკითხთან დაკავშირებით საკუთარი აზრის, პოზიციის გამოსატვა**, 2) **გამოვლენა, აღნუსხვა; აღბეჭდვა და** 3) **მიმაგრება/ დამაგრება**. აღსანიშნავია, რომ ამ მნიშვნელობებს დართული აქვს საკმაოდ ვრცელი საილუსტრაციო მასალა, რომელიც ქართული ენის ვებკორპუსიდანაა ამოღებული.

## 12. **ექსკლუზივიზი**

### **ქართული ენის განმარტებითი ლექსიკონი**

**ექსკლუზივიზი-0 (ექსკლუზივიზისა)** [ფრან. exclusive გამომრიცხველი] *გრამატ.* პირველი პირი მრავლ. რიცხვისა, რომელიც შეიცავს პირველ და მესამე პირს და გამომრიცხავს მეორეს; ამჟამად დამახასიათებელია სვანური ენისა (ნიშგვეიჩემი და მისი, ჩემი და მათი; ნალატგვიყვარს მე და მას, მე და მათ) (საპირისპ. ინკლუზივი).

აღნიშნული სიტყვა არ არის მოცემული „ქართულ ლექსიკონში“.

### **ქართული ენის ვებკორპუსი**

სიხშირული ინდექსი: 19026

**ექსკლუზივიზი-0 (ექსკლუზივიზისა)** 1. **ამა თუ იმ ტელევიზიის, რადიოს, ჟურნალის და მისთ. მიერ უახლესი ინფორმაციის ფლობა და გავრცელება; ელექტრონულ ფოსტას გადაგავლე თვალი და ექსკლუზივებზე დავიწყე ნადირობა; ექსკლუზივი მართლაც გამოგვივიდა, რადგან ძეგლის აღების ფოტოების ფოტო და ვიდეომასალის გადაღება მხოლოდ ჩვენ მოვახერხეთ; არც**



„ნანუკას შოუს“ ექნება მსგავსი ექსკლუზივი; **2. რისამე პირველად გამოყენების უფლება;** გავრცელებული ინფორმაციით, მოპოვების ტექნოლოგიებზე ექსკლუზივი ამერიკულ კომპანიებს ეკუთვნით; ყველა ტენდერ-პროექტსა და სახელმწიფო შეკვეთა-მიწოდებების უზრუნველყოფაზე ექსკლუზივი ჰქონდათ; **3. გრამტ. პირველი პირი მრავლ. რიცხვისა, რომელიც შეიცავს პირველ და მესამე პირს და გამორიცხავს მეორეს;** სვანური ენის რიგი სპეციფიკური საკითხი (ბრუნების ორფუძიანობა, ექსკლუზივ-ინკლუზივის პრობლემა.

ქართული ენის განამრტებით ლექსიკონში სალექსიკონო ერთეულისათვის **ექსკლუზივი** გამოტანილია მხოლოდ ერთადერთი მნიშვნელობა, რომელიც ლინგვისტური ტერმინია, პირველი პირი მრავლ. რიცხვისა, რომელიც შეიცავს პირველ და მესამე პირს და გამორიცხავს მეორეს; ქართული ენის ვებკორპუსის მიხედვით შედგენილ სალექსიკონო სიტყვა-სტატიაში კი მოცემული სიტყვის სამი მნიშვნელობა გამოიყოფა. პირველ ადგილზე გამოტანილია ისეთი მნიშვნელობა, რომელიც დღევანდელ ქართულ მეტყველებაში ძალიან აქტიურად გამოიყენება, კერძოდ, **ამა თუ იმ ტელევიზიის, რადიოს, ჟურნალის და მისთ. მიერ უახლესი ინფორმაციის ფლობა და გავრცელება;** ელექტრონულ ფოსტას გადავავლე თვალი და ექსკლუზივებზე დავიწყე ნადირობა; ექსკლუზივი მართლაც გამოგვივიდა, რადგან ძეგლის აღების ფოტოების ფოტო და ვიდეომასალის გადაღება მხოლოდ ჩვენ მოვახერხეთ; არც „ნანუკას შოუს“ ექნება მსგავსი ექსკლუზივი. არანაკლებ საინტერესოა ამ სიტყვის მეორე მნიშვნელობა, რომელიც ასევე აქტიურად გამოიყენება თანამედროვე ქართულ ენაში **რისამე პირველად გამოყენების უფლება;** გავრცელებული ინფორმაციით, მოპოვების ტექნოლოგიებზე ექსკლუზივი ამერიკულ კომპანიებს ეკუთვნით; ყველა ტენდერ-პროექტსა და სახელმწიფო შეკვეთა-მიწოდებების უზრუნველყოფაზე ექსკლუზივი ჰქონდათ. ამგვარად, ვფიქრობთ რომ მოცემული სიტყვის ხემოთ წარმოდგენილი მნიშვნელობები ქართული ენის ვებკორპუსის მიხედვით, წარმოადგენს ახალ რეალიას, ახალ მნიშვნელობას ამ სიტყვისათვის, რომლის ასახვა აუცილებელია ქართული ენის განმარტებით ლექსიკონში. რაც შეეხება „ქართულ ლექსიკონს“, აღნიშნული სიტყვა არაა მოცემული.

### 13. ევმატური

#### ქართული ენის განმარტებითი ლექსიკონი

**ევმატური-ი (ევმატურისა) წიგნ. ევმატის (მნიშვ. 2) მომცემი, –** სასურველი შედეგის მომცემი, ქმედითი, შედეგიანი. *ვეელაფერი უნდა გავაკეთოთ იმისათვის, ...რომ აღძრულ საკითხზე მიღებულ იქნეს ქმედითი და ევმატური ზომები («კომ.»).* **ევმატურად ზმნს.** მოქმედად, შედეგიანად. *განსაკუთრებული ყურადღება ექცევა მინერალური სასუქების ევმატურად გამოყენებას («კომ.»).*

#### „ქართული ლექსიკონი“

**ევმატური ევმატურის 1. დიდი ევმატის მქონე. ევმატური ღონისძიება. ევმატური გამოსვლა. 2. ევმატის მომცემი. ევმატური საშუალება. ევმატური მედიკამენტები.**

#### ქართული ენის ვებკორპუსი

სისშირული ინდექსი: 11838

**ევმატური-ი (ევმატურისა) 1. ევმატის მომცემი, სასურველი შედეგის მომცემი, ქმედითი, შედეგიანი; შეწყვიტეთ დიეტა, რადგან ევმატური არ იქნება და ჯანმრთელობასაც გაიუარესებთ; ამ მუავის მოქმედება საკმაოდ ევმატური აღმოჩნდა მწვავე რესპირატორული დაავადებების გამოსაკვლევად ჩატარებული ცდებისას; გუნდის ძლიერი და სუსტი მხარეების ევმატური გამოყენება და როლების შესაბამისი გადანაწილება; დანაშაულის ევმატური გამოძიება და სისხლისსამართლებრივი დევნა ტრეფიკინგში მონაწილე პირებისადმი; 2. ლამაზი, მიმზიდველი; საშა პრიშინის მიერ გადაღებული მოდელების ფოტოები ევმატური გამოვიდა; შეიძლება გეცვას სადა კაბა და ამ დროს გეკეთოს ერთი გამორჩეული სამკაული, ან თმის სამაგრი, რომ თვალში საცემი და ევმატური გახდეს; ჯინსის ფრჩხილებიც არანაკლებ ევმატური და ორიგინალურია.**

ქართული ენის განმარტებით ლექსიკონში სალექსიკონო ერთეული **ევმატური** წარმოდგენილია ერთადერთი მნიშვნელობით - **ევმატის მომცემი, სასურველი შედეგის მომცემი, ქმედითი, შედეგიანი.** ხოლო „ქართულ ლექსიკონში“ გამოყოფილია ამავე სიტყვის ორი მნიშვნელობა. რაც შეეხება ქართული ენის ვებკორპუსის მიხედვით შედგენილ სიტყვა-სტატიას, მასში ასახულია ამ სიტყვის ისეთი მნიშვნელობა, რომელიც არაა მოცემული

ზემოაღნიშნულ ლექსიკონებში, კერძოდ, **ლამაზი, მიმზიდველი; საშა პრიშვინის მიერ გადაღებული მოდელების ფოტოები ეფექტური გამოვიდა; შეიძლება გეცვას სადა კაბა და ამ დროს გეკეთოს ერთი გამორჩეული სამკაული, რომ თვალში საცემი და ეფექტური გახდეს; შეიძლება, ყვითელი პალტო არ იყოს მოდური, მაგრამ იმდენად ეფექტურმა ქალმა ჩაიცვას, რომ ის სამოსი აქტუალურად გეჩვენოს.**

#### 14. მღვრადი

##### ქართული ენის განმარტებითი ლექსიკონი

**მღვრად-ი (მღვრადისა) სპეც. გარკვეული მდგომარეობის შენარჩუნების უნარის მქონე** (მიუხედავად სხვადასხვა ძალთა მოქმედებისა). მღვრადი წონასწორობა. – მღვრადი მდგომარეობა || გამძლე.

აღნიშნული მნიშვნელობა არ არის მოცემულ „ქართულ ლექსიკონში“

##### ქართული ენის ვებკორპუსი

სისშირული ინდექსი: 9280

**მღვრად-ი (მღვრადისა) 1. გამძლე;** განსაკუთრებულად მძაფრი და მღვრადი სურნელი აქვს; არსებობს სპეციალისტთა დასკვნები, რომ ეს მღვრადი შენობა იყო და არავის უქმნიდა საფრთხეს; **2. გაუტეხელი, ძლიერი;** იგი ძლიერი და მღვრადი ადამიანია; ფიზიკური მონაცემები და ვითომ მღვრადი ხასიათი ნუ მოგატყუებს. ამ ყველაფრის უკან ერთი სუსტი გოგო დგას; შინაგანად, ძალიან მღვრადი კაცი იყო, შეეძლო ზრდილობიანად გასცლოდა იმ სიტუაციას, რომელიც მიუღებელი იყო მისთვის; **3. სტაბილური;** ეკონომიკის ზრდა მღვრადი ვერ იქნება, თუ მაკროეკონომიკური გარემო სტაბილური არ იქნა; ლიდერების მიღწევები იმით იზომება, რამდენად მღვრადი დარჩება ისინი მას შემდეგ, რაც თავად ძალაუფლებიდან წავლენ; **4. სპეც. გარკვეული მდგომარეობის შენარჩუნების უნარის მქონე;** რამდენიმე ქიმიური რეაქციის შემდეგ წარმოიქმნება უსუნო, პლასტიკური, თერმულად მღვრადი, ნეიტრალური ნივთიერება; დიოქსინები მღვრადი ნაერთებია, არ იშლება, ინტენსიურად გროვდება როგორც ბიოსფეროში, ასევე ადამიანის ორგანიზმშიც; **მღვრადი ბანკოთარება;** აუქციონში გამარჯვებული ინვესტორის ვინაობის შესახებ

ინფორმაციას ეკონომიკისა და მდგრადი განვითარების სამინისტრო ვებ-გვერდზე მოგვიანებით განათავსებს; სტრატეგიის მიზანი დარგის მდგრადი განვითარების ხელშემწყობი გარემოს შექმნა; ქალაქის მდგრადი განვითარებისთვის სხვადასხვა ქვეყანაში სხვადასხვა მიდგომები აქვთ.

ქართული ენის განამრტებით ლექსიკონში სალექსიკონო ერთეული **მდგრადი წარმოდგენილია ერთი მნიშვნელობით - გარკვეული მდგომარეობის შენარჩუნების უნარის მქონე.** ქართული ენის ვებკორპუსის მიხედვით შედგენილ სალექსიკონო სიტყვა-სტატიაში მოცემული სიტყვის ოთხი მნიშვნელობა გამოვყავით. კერძოდ, 1) **გამძლე; განსაკუთრებულად მძაფრი და მდგრადი სურნელი აქვს,** 2) **გაუტყეული, ძლიერი; იგი ძლიერი და მდგრადი ადამიანია;** 3) **სტაბილური; ეკონომიკის ზრდა მდგრადი ვერ იქნება, თუ მაკროეკონომიკური გარემო სტაბილური არ იქნა.** აღსანიშნავია, რომ ბოლო მეოთხე მნიშვნელობად არის გამოტანილი ქართული ენის განმარტებით ლექსიკონში მოცემული მნიშვნელობა - **გარკვეული მდგომარეობის შენარჩუნების უნარის მქონე; რამდენიმე ქიმიური რეაქციის შემდეგ წარმოიქმნება უსუნო, პლასტიკური, თერმულად მდგრადი, ნეიტრალური ნივთიერება.** გარდა ამ მნიშვნელობებისა, სალექსიკონო სიტყვა-სტატიაშია მოცემულია ისეთი შესიტყვება, როგორცაა **მდგრადი განვითარება,** რაც არ არის ასახული ქართული ენის განმარტებით ლექსიკონში. რაც შეეხება „ქართულ ლექსიკონს“, აღნიშნული სიტყვა არაა მოცემული.

## 15. კომპლემსი

### ქართული ენის განმარტებითი ლექსიკონი

**კომპლემს-ი (კომპლემსისა)** [ლათ. complexus წნული, ხლართი] 1. ერთგვაროვან საგანთა ან მოვლენათა ერთობლიობა, რომელიც მთელს, მთლიანს რასმე წარმოადგენს. მაღარიის საწინააღმდეგო ღონისძიებათა კომპლექსი. – თათბირზე განიხილეს პოლიტექნიკურ სწავლებასთან დაკავშირებულ საკითხთა კომპლექსი. 2. არქიტ. ნაგებობათა წყება, რომელთაც საერთო დანიშნულება აქვთ. პოლიტექნიკური ინსტიტუტის სასწავლო და საცხოვრებელ შენობათა კომპლექსი («ლიტ. და ხელოვნ»). ენერგო-ირიგაციულ ნაგებობათა ეს უდიდესი კომპლექსი იქნება ერთ-ერთი უმნიშვნელოვანესი რგოლი ბუნების გარდაქმნის გრანდიოზული სტალინური გეგმის ღონისძიებათა წყებაში («ახ. კომ.»). 3. ლინგვ. რამდენიმე ბგერისაგან შედგენილი ჯგუფი. ქართული ენისთვის თანხმოვანთა ბუნებრივი კომპლექსებია: ცხ, წყ, ტკ... – თვით «ვობი»-სა და «ვიბი»-ს კომპლექსებშიც არაფერი ისეთი არ მოიპოვება, რომელიც ამ სახელის სამწერლო ქართულში ყოფნას შეუძლებლად ხდიდეს (ივ. ჯავახ.).

### „ქართული ლექსიკონი“

**კომპლემსი** კომპლექსის საგნების, მოვლენების, თვისებების ერთობლიობა, რომელიც გარკვეულ მთლიანობას ქმნის. შენობების კომპლექსი.

### ქართული ენის ვებკორპუსი

სისშირული ინდექსი: 6292

**კომპლემს-ი (კომპლემსისა)** 1. რისამე დაგმობა/შიში ამა თუ იმ ტრადიციის, შეხედულებისა და მისთ. გამო; მუდმივად მაქვს იმის კომპლექსი, რომ რაღაცას სწორად არ ვაკეთებ; ჩემს შვილს ადამიანებთან ურთიერთობის კომპლექსი აქვს, კარგია, დროზე მიეხვევა თანატოლებთან კონტაქტს; გულს მიკლავს ჩემი აღნაგობა და ამის გამო უამრავი რამის კომპლექსი გამიჩნდა; 2. ერთგვაროვან საგანთა ან მოვლენათა ერთობლიობა, რომელიც მთელს, მთლიანს რასმე წარმოადგენს; ჯანსაღ კვებაზე გადაყვანა და ვიტამინების კომპლექსი ბავშვებს მდგომარეობიდან გამოსვლაში დაეხმარება; მაგალითად, ძალიან კარგია კვების ვარჯიშთა კომპლექსი; 3. არქიტ. ნაგებობათა წყება, რომელთაც საერთო დანიშნულება აქვთ; არმაზისხევში არის შუა საუკუნეების

სამონასტრო კომპლექსი; თამარაშვილის ქუჩაზე მშენებარე საცხოვრებელი კომპლექსი დამოუკიდებელი 9 შენობისაგან შედგება; **არასრულფასოვნების კომპლექსი**; ეჭვიანობის მიზეზად ხშირად სახელდება არასრულფასოვნების კომპლექსი, თავმოყვარეობის შელახვის ან საყვარელი ადამიანის დაკარგვის შიში; მიხნეულია, რომ გოგონებს უფრო ხშირად ტანჯავთ არასრულფასოვნების კომპლექსი; **ოიდიპოსის კომპლექსი**; **ოიდიპოსის კომპლექსი** და **ელექტრას კომპლექსი** ფსიქოსექსუალური განვითარების მესამე საფეხურზე იჩენს თავს; **ოიდიპოსის კომპლექსი** არის ფსიქოანალიტიკური წარმოდგენა 3-4 წლის ასაკის ბავშვის მშობლებისადმი ემოციონალური დამოკიდებულებების თავისებურებების შესახებ.

ქართული ენის განმარტებით ლექსიკონში სალექსიკონო ერთეული **კომპლექსი** წარმოდგენილია სამი მნიშვნელობით. თუმცა, მასში ასახვა ვერ პოვა ამ სიტყვის ისეთმა გავრცელებულმა მნიშვნელობამ, როგორცაა **რისამე დაგმობა/შიში ამა თუ იმ ტრადიციის, შეხედულებისა და მისთ. გამო**; მუდმივად მაქვს იმის კომპლექსი, რომ რაღაცას სწორად არ ვაკეთებ; ჩემს შვილს ადამიანებთან ურთიერთობის კომპლექსი აქვს, კარგია, დროზე მიეჩვევა თანატოლებთან კონტაქტს; გულს მიკლავს ჩემი აღნაგობა და ამის გამო უამრავი რამის კომპლექსი გამიჩნდა. გარდა ამ მნიშვნელობისა, წარმოდგენილია მოცემულია ისეთი შესიტყვებები როგორცაა, **არასრულფასოვნების კომპლექსი**; **ეჭვიანობის მიზეზად ხშირად სახელდება არასრულფასოვნების კომპლექსი**, **ოიდიპოსის კომპლექსი**; **ოიდიპოსის კომპლექსი** და **ელექტრას კომპლექსი** ფსიქოსექსუალური განვითარების მესამე საფეხურზე იჩენს თავს. რაც შეეხება „ქართულ ლექსიკონს“, აღნიშნული სიტყვის მხოლოდ ერთი მნიშვნელობაა მოცემული.

## 16. გეი

### ქართული ენის განმარტებითი ლექსიკონი

**გეი (გეისა)** [ინგ. gay] ბარბარ. მამათმავალი. ბერლინის სკოლაში გეებთან ურთიერთობის გაკვეთილები შემოიღეს («რეიტა»).

აღნიშნული სიტყვა არაა მოცემული „ქართულ ლექსიკონში“.

### ქართული ენის ვებკორპუსი

სიხშირული ინდექსი: 5975

**გეი (გეისა)** – მამათმავალი; ზუსტად ვიცოდი, რომ ერთი ბიჭი „გეი“ იყო, მაგრამ მასაც ვერ გავუბედე ამ თემაზე ლაპარაკი, თუ ვინმე არის გეი, მაგრამ უფალს ეძებს და მას კეთილი განზრახვები აქვს, ვინ ვარ მე, რომ ის განვსაჯო? **გეი აღლუმი**; ფაქტია, რომ გეი აღლუმად შერაცხული ერთეული ადამიანების მცირემარშრუტიანი მსვლელობა, ვერ მიიქცევდა ესოდენ დიდ ყურადღებას, ეს კანონი, ფაქტობრივად, კრძალავს გეი აღლუმებს; **გეი პარადი**; გეი პარადის ჩატარება აკრძალა სერბეთის ხელისუფლებამ ბელგრადში; თბილისში აპირებდნენ გეი პარადს და მაშინ მთელი თბილისი ემზადებოდა ამის დასარბევად.

ქართული ენის განმარტებით ლექსიკონში სალექსიკონო ერთეული **გეის** მითითებული აქვს სტილისტური კვალიფიკაცია – ბარბარიზმი. რაც შეეხება ქართული ენის ვებკორპუსის მიხედვით შედგენილ სალექსიკონო სიტყვა-სტატიაში საილუსტრაციო მასალასთან ერთად ასახულია ამ სიტყვასთან დაკავშირებული ისეთი შესიტყვებები როგორცაა, **გეი აღლუმი**, **გეი პარადი** და ა.შ. რაც შეეხება „ქართულ ლექსიკონს“, აღნიშნული სიტყვა არაა მოცემული.

## 17. სიუჟეტი

### ქართული ენის განმარტებითი ლექსიკონი

**სიუჟეტი-0 (სიუჟეტისა)** [ფრან. sujet საგანი] მხატვრულ ნაწარმოებში – ერთმანეთთან დაკავშირებული მოქმედებანი და ამბები, რომლებიც გამოავლენენ მოქმედ პირთა ხასიათსა და ურთიერთობას. დრამატული სიუჟეტი საინტერესო და მოძრავი ვერ იქნება, თუ მას საფუძველად რაიმე მნიშვნელოვანი ხასიათის კონფლიქტი არ უძევს (გ. ქიქ.). ბალეტ «გორდას» სიუჟეტი ქართული ლეგენდის «სურამის ციხის» მოტივს შეიცავს («კომ»).

### „ქართული ლექსიკონი“

**სიუჟეტი** სიუჟეტის, ლიტერატურულ ნაწარმოებში გადმოცემული ფაქტების, მოვლენების თანმიმდევრობა. რომანის სიუჟეტი. აღმოსავლური სიუჟეტი. სიუჟეტური. სიუჟეტთან დაკავშირებული. სიუჟეტისთვის დამახასიათებელი. სიუჟეტური ხაზი.

### ქართული ენის ვებკორპუსი

სიხშირული ინდექსი: 5421

**სიუჟეტი-0 (სიუჟეტისა)** 1. სატელევიზიო, რადიო გადაცემებში სპეციალურად მომზადებული ეპიზოდი; 15 სექტემბერს ტელევიზიის ეთერით გავიდა სიუჟეტი, რომელიც ეხებოდა საქართველოს სარაგბო ჩემპიონატის უმაღლეს ლიგაში მოასპარეზე კლუბ, „არმაზის“ ორ მორაგბეს, რამდენიმე წუთის წინ ტელეკომპანია „იმედის“ ეთერით გავიდა სიუჟეტი, რომელიც ვერანაირ კრიტიკას ვერ უძლებს; საკონკურსოდ არ შემოსულა არც ერთი რადიო სიუჟეტი; 2. მხატვრულ ნაწარმოებში ერთმანეთთან დაკავშირებული მოქმედებანი და ამბები, რომლებიც გამოავლენენ მოქმედ პირთა ხასიათსა და ურთიერთობას; ის ყველა ფილმს აერთიანებს ერთი სიუჟეტი - ქალები და კაცები ცდილობენ თავიანთი დროის და თავიანთი არსებობის სივრცის ორგანიზებას; ისეთი შთაბეჭდილება დამრჩა, რომ თითქონს ნაწყვეტია ნაწარმოებიდან და არა მთლიანი სიუჟეტი.

ქართული ენის განმარტებით ლექსიკონისა და „ქართული ლექსიკონის“ სალექსიკონო ერთეულის **სიუჟეტის** მხოლოდ ერთი მნიშვნელობაა მოცემული, რომელიც უკავშირდება მხატვრულ ნაწარმოებში ერთმანეთთან დაკავშირებულ



მოქმედებებსა და ამბებს. რაც შეეხება ქართული ენის ვებკორპუსის მიხედვით შედგენილ სალექსიკონო სიტყვა-სტატიას, აქ პირველ მნიშვნელობად გამოტანილია შემდეგი: **სატელევიზიო, რადიო გადაცემებში სპეციალურად მომზადებული ეპიზოდი; 15 სექტემბერს ტელევიზიის ეთერით გავიდა სიუჟეტი, რომელიც ეხებოდა საქართველოს სარაგბო ჩემპიონატის უმაღლეს ლიგაში მოასპარეზე კლუბ, „არმაზის“ ორ მორაგბეს; რამდენიმე წუთის წინ ტელეკომპანია „იმედის“ ეთერით გავიდა სიუჟეტი, რომელიც ვერანაირ კრიტიკას ვერ უძლებს; საკონკურსოდ არ შემოსულა არც ერთი რადიო სიუჟეტი.** აღსანიშნავია, რომ ზემოთ მოცემული წარმოადგენს ამ სიტყვის ყველაზე მეტად გავრცელებულ მნიშვნელობას დღევანდელ ქართულ ენაში, განსაკუთრებით მედია საშუალებებში.

## 18. მობილური

### ქართული ენის განმარტებითი ლექსიკონი

**მობილურ-ი (იმობილურისა),** იმობილიზებული-ი (იმობილიზებულისა) მედიც. რასაც იმობილიზაცია უყვეს.

**მობილურ-ი (იმობილურისა) სწრაფი მოძრაობის უნარის მქონე, მოძრავი.** ჯარის. მობილური ნაწილი – მთელ დასავლეთ გერმანიაში იქმნება მობილური პოლიცია 30 ათასი კაცის რაოდენობით («კომ.»).

### „ქართული ლექსიკონი“

**მობილური მობილურის რაც ადვილად გადაადგილდება, მოძრავი.** მობილური დანადგარი. მობილური ტელეფონი.

### ქართული ენის ვებკორპუსი

სიხშირული ინდექსი: 5030

**მობილურ-ი (იმობილურისა) არსებითი სახელი; იგივეა რაც მობილური ტელეფონი; დღეს ყველას აქვს მობილური, დღევანდელი ტექნიკა საუკუნეების წინ ზღაპარი ეგონებოდათ; პაატამ მობილური ამოიღო და სადღაც დარეკა; დაფიქრდით იმაზეც, ხომ არ გაუფანტავს მობილური ბავშვს ყურადღებას გაკვეთილებზე.**

**მოზილურ-ი (მოზილურისა) ზედსართავი სახელი, სწრაფი მოძრაობის უნარის მქონე, მოძრავე; მობილური, ბრძოლისუნარიანი, პირველხარისხოვანი რეგულარული არმია; გაიოლდა მიმოსვლა, ადამიანი ვახდა უფრო მობილური; მობილური აპლიკაცია; 2013 წლის მაისში შეიქმნა მობილური აპლიკაციის სერვისი; ვიდეოთვალი აქვს ჩამონტაჟებული და მრავალი მობილური აპლიკაციის დაყენების საშუალებას იძლევა; მობილური ინტერნეტი; საქართველოში მობილური ინტერნეტი ნაკლებად განვითარებულია; ზოგიერთ მოდელს კი 3G მობილური ინტერნეტის დაჭერაც შეუძლია; მობილური ოპერატორი; მთელი საჯარო სამსახური დაავადებულეს მხოლოდ ერთი მობილური ოპერატორით ესარგებლა; ამ სერვისით სარგებლობა ნებისმიერი მობილური ოპერატორის აბონენტს შეუძლია; მობილური ტელეფონი; განვითარებად ქვეყნებში მობილური ტელეფონების ათვისების მზარდი ტემპი შეინიშნება; არსებობს მობილური ტელეფონი, რომლის მეშვეობითაც შეგიძლია გამოხატო საყვარელი ადამიანის მიმართ ყურადღება; ქალმა კი ჩანთიდან მობილური ტელეფონი ამოიღო და სადღაც დარეკა.**

ქართული ენის განმარტებით ლექსიკონში სალექსიკონო ერთეული **მობილური** წარმოდგენილია ორი მნიშვნელობით. პირველი სამედიცინო მნიშვნელობით, **რასაც იმობილიზაცია უყვეს**, მეორე მნიშვნელობაა: **სწრაფი მოძრაობის უნარის მქონე, მოძრავე.** „ქართულ ლექსიკონში“ კი ამ სიტყვის ერთი მნიშვნელობაა მოცემული. ქართული ენის ვებკორპუსის მიხედვით შედგნილ სალექსიკონო სიტყვა-სტატიაში ცალკე ერთეულად არის გამოტანილი როგორც არსებითი სახელი: **იგივეა რაც მობილური ტელეფონი; დღეს ყველას აქვს მობილური, პატამ მობილური ამოიღო და სადღაც დარეკა.** რაც შეეხება მეორე სიტყვა-სტატიას, როგორც ზედსართავ სახელს, აქ წარმოდგენილია იმ მნიშვნელობით, რა მნიშვნელობითაც, ეს ქართული ენის განმარტებით ლექსიკონშია მოცემული. თუმცა, მოცემულია ისეთი შესიტყვებები, რაც არ არის ასახული ქართული ენის განმარტებით ლექსიკონში, მაგალითად: **მობილური აპლიკაცია; 2013 წლის მაისში შეიქმნა მობილური აპლიკაციის სერვისი; მობილური ინტერნეტი; საქართველოში მობილური ინტერნეტი ნაკლებად განვითარებულია; მობილური ოპერატორი; ამ სერვისით სარგებლობა ნებისმიერი მობილური ოპერატორის აბონენტს შეუძლია; მობილური ტელეფონი; განვითარებად ქვეყნებში მობილური ტელეფონების ათვისების მზარდი ტემპი შეინიშნება.**

## 19. დერეფანი

### ქართული ენის განმარტებითი ლექსიკონი

**დერეფან-ი** (დერეფნისა) 1. სახლის ნაწილი, წინა მხარეს მიდგმული შესასვლელად, უკედლებო (ღია), სვეტებიანი და გადახურული (სინონ. წინკარი, საჩხეხი). ბუდეს გააკეთებს შენი დერეფნის ჭერში მერცხალი (თ. რაზიკ). [ბაბუცამ] მარდად დაგავა... როგორც სახლი, ისე დერეფანი და კარ-მიდამო (ნ. ლომ). რაც შეეხება სახლს, ეს იყო... ძველი ფიცრული შენობა, ...მი წის იატაკით და წინ მიშენებული დერეფნით (ე. ნინოშ). ვიდრე [ბერი] გულს მოიჯერებდა მონატრებული დედამამისშვილის ალერსით, ეზო და სასახლის ფართო დერეფანი შინამოსამსახურებით აივსო (რ. ჯაფ.). იმ დამეს გიორგამ დერეფანში მოითხოვა დაძინება, – გარეთ, ტახტზე გამიშაღეთო ლოგინი (ნ. ონაშ). დერეფნის ჩიტი კუთხ. (ქართლ.) მერცხალი. 2. გრძელი გასავლელი (სადგომში...), – ტალანი, კორიდორი. გავიარეთ ბნელი, ვიწრო დერეფანი და შევედით ერთ ოთახში («ცისკარი»). რაიონის საავადმყოფოს დერეფანში ტარასი... ექიმს ელოდებოდა (ს. კლდ.). ვიწრო დერეფანი ბევრად უფრო პატარა ორ სხვა ოთახს აკავშირებდა (მ. ბაქრ. თარგმ.). სკოლაში ადრე მივედი, დირექტორს დერეფანში დაველოდე (ვ. ჩუკურ.).

### „ქართული ლექსიკონი“

**დერეფანი** დერეფნის 1. შენობის შიდა ნაწილში მოქცეული გასასვლელი, რომელიც ოთახებიდან არის დაკავშირებული. დერეფნის კარი დარბაზში გადიოდა. 2. საცხოვრებელი სახლის წინა მხარეს მიდგმული, სვეტებზე დაყრდნობილი გადახურული ნაწილი. ღია დერეფანი.

### ქართული ენის ვებკორპუსი

სიხშირული ინდექსი: 1925

**დერეფან-ი** (დერეფნისა) 1. სახლის ნაწილი, წინა მხარეს მიდგმული შესასვლელად, უკედლებო (ღია), სვეტებიანი და გადახურული; წამოხტა, საძინებლიდან დერეფანი გაირბინა და საერთო ოთახში ბუხრის წინ მოკალათდა; პატარა სახლი გვექონდა, ერთი ოთახი და ერთი პატარა დერეფანი, 2. გრძელი გასავლელი, კორიდორი; ექთანმა სწრაფად გაიარა დერეფანი, რიგში მდგომებს მიუვლო, ვინც შეხვევაზე ხართ, შემოდითო და კაბინეტში გაუჩინარდა; კარგა

გრძელი იყო დერეფანი, ბოლოს კიბით მთავრდებოდა; **3. ამა თუ იმ ქვეყანასთან დამაკავშირებელი სავაჭრო, სატრანზიტო და ა.შ. გზა;** საქართველოზე გადის ენერგეტიკული დერეფანი, როგორც აღმოსავლეთიდან დასავლეთისკენ, ასევე ჩრდილოეთიდან სამხრეთისკენ; ბაქო-თბილისი-ყარსის რკინიგზა გახდება ევროპა-კავკასია-აზიის უმნიშვნელოვანესი სარკინიგზო დერეფანი; **4. საგანგებლო გასასვლელი;** პოლიციას ჰქონდა ჯებირებით გაკეთებული სამი დერეფანი; თავდაცვის წარმომადგენლებს მათთვის დერეფანი აქვთ გაკეთებული და ცდილობენ სამხედროები უვნებლად გაატარონ; **სირცხვილის დერეფანი;** საშინელი რეაქციით შეგვხვდნენ, სირცხვილის დერეფანი მოგვიწყვეს; ვაპირებდით, კანონის ფარგლებში მოგვეწყო სირცხვილის დერეფანი; **სატრანსპორტო დერეფანი;** რუსეთისა და ცენტრალური აზიის ქვეყნებთან დამაკავშირებელი სატრანსპორტო დერეფანი; აღმოსავლეთ-დასავლეთის სატრანსპორტო დერეფანი ქვეყნისთვის მნიშვნელოვანი და სტრატეგიული სავაჭრო არტერიაა; **სატრანზიტო დერეფანი;** სამხრეთ კავკასია და შესაბამისად საქართველო, არის ამ რეგიონზე გამავალი სატრანზიტო დერეფანი; სატრანზიტო დერეფანი პირდაპირ დაკავშირებულია საქართველოს განვითარების პერსპექტივებთან; **სავაჭრო დერეფანი;** შეთანხმების მიხედვით, შეიქმნება სამი სავაჭრო დერეფანი - ფსოუ-ზუგდიდის, როკი-გორისა და ლარსის; მესამე სავაჭრო დერეფანი ბევრად მცირე იქნება და მონიტორების ორი ჯგუფი საზღვრის ორივე მხარეს განლაგდება.

ქართული ენის განმარტებით ლექსიკონსა და „ქართულ ლექსიკონში“ სალექსიკონო სიტყვა-სტატია **დერეფანი** წარმოდგენილია ორი მნიშვნელობით. რაც შეეხება ქართული ენის ვებკორპუსის მიხედვით შედგნილ სიტყვა-სტატიას, მოცემული სიტყვა წამოდგენილია ოთხი მნიშვნელობით, მათ შორის ისეთი მნიშვნელობით, რაც არ არის ასახული ქართული ენის განმარტებით ლექსიკონში. მაგალითად, მესამე მნიშვნელობად გამოვიტანეთ: **ამა თუ იმ ქვეყანასთან დამაკავშირებელი სავაჭრო, სატრანზიტო და ა.შ. გზა;** საქართველოზე გადის ენერგეტიკული დერეფანი, როგორც აღმოსავლეთიდან დასავლეთისკენ, ასევე ჩრდილოეთიდან სამხრეთისკენ; ბაქო-თბილისი-ყარსის რკინიგზა გახდება ევროპა-კავკასია-აზიის უმნიშვნელოვანესი სარკინიგზო დერეფანი. მეოთხე მნიშვნელობად კი გამოტანილია: **საგანგებლო გასასვლელი;** პოლიციას ჰქონდა ჯებირებით გაკეთებული სამი დერეფანი; თავდაცვის წარმომადგენლებს მათთვის დერეფანი აქვთ გაკეთებული და ცდილობენ სამხედროები უვნებლად გაატარონ. აღსანიშნავია, რომ ზემოთ წარმოდგენილი

ამ სიტყვის ერთ-ერთი ყველაზე გავრცელებული მნიშვილობებია და მისი ასახვა განამრტებით ლექსიკონში აუცილებლობას წარმოადგენს. გარდა ამ მნიშვილობებისა, სალექსიკონო სიტყვა სტატიაში ასახვა პოვა ისეთმა შესიტყვებებმა, როგორცაა: **სირცხვილის დერეფანი**; *საშინელი რეაქციით შეგვხვდნენ, სირცხვილის დერეფანი მოგვიწყვეს*; **სატრანსპორტო დერეფანი**; *რუსეთისა და ცენტრალური აზიის ქვეყნებთან დამაკავშირებელი სატრანსპორტო დერეფანი*; **სატრანზიტო დერეფანი**; *სატრანზიტო დერეფანი პირდაპირ დაკავშირებულია საქართველოს განვითარების პერსპექტივებთან*; **სავაჭრო დერეფანი**; *შეთანხმების მიხედვით, შეიქმნება სამი სავაჭრო დერეფანი - ფსოუ-ზუგდიდის, როკი-გორისა და ღარსის.*

## 20. კორექტირება

### ქართული ენის განმარტებითი ლექსიკონი

კორექტირება (კორექტირებისა) წიგნ. კორექტივის, შესწორების შეტანა.

„ქართული ლექსიკონი“

კორექტირება კორექტირების შესწორება, კორექტივის შეტანა. ნაშრომის კორექტირება.

### ქართული ენის ვებკორპუსი

სისშირული ინდექსი: 1605

კორექტირება (კორექტირებისა) კორექტივის, შესწორების შეტანა; *მოცემული დასახლებებისათვის ბალებში დადგენილი სეისმურობის სიდიდის კორექტირება*; *ფინანსთა სამინისტროს რეალურად მოუწვეს სახელმწიფო ბიუჯეტის კორექტირება*; *აუცილებელია მოხდეს ხარვეზების კორექტირება და ზარალის კომპენსაცია.*

## 21. ბენეფიციარი

### ქართული ენის განმარტებითი ლექსიკონი

**ბენეფიციარი-ი (ბენეფიციარი-ისა)** ნეოლოგ. გენური ინჟინერიის შედეგი, გენშეცვლილი. პერპეტუალური გენმოდულირებული ვირუსის ბაზაზე დამზადებული პრეპარატის კლინიკური გამოცდა უკვე დაიწყო («ას.-დას»). არ არსებობს მავნებელი გენმოდულირებული პროდუქტების ნუსხა («კვ. პალიტ.»).

აღნიშნული სიტყვა არაა მოცემული „ქართულ ლექსიკონში“.

### ქართული ენის ვებკორპუსი

სისშირული ინდექსი: 1194

**ბენეფიციარი-ი (ბენეფიციარი-ისა)** 1. გენური ინჟინერიის შედეგი, გენშეცვლილი; კარტოფილი შეიძლება იყოს გენმოდულირებული, ნიტრატებისა და პესტიციდების შემცველი; ოსაკას უნივერსიტეტის მეცნიერებმა შექმნეს გენმოდულირებული თაგვი, რომელსაც ჩიტვით ელურტული შეუძლია; ევროპული უნივერსიტეტის ექსპერტებმა შეიმუშავეს გენმოდულირებული თამბაქოს მცენარე, რომელიც მაღარი რეზისტენტულ ფორმას ებრძვის; 2. გადატ. მდარე, უხარისხო; ახალ ამბლუაში მოგვევლინა თავისი მორიგი გენმოდულირებული ლექსით; ეს ყბადაღებული გენმოდულირებული პოეზია ვის არ მოსწონს; ზედმეტად მორცხვი, გომურად კეკლუცი და საშუალოზე ღამაზი, გენმოდულირებული ადამიანური თვისებების საცავს წარმოადგენდა; **ბენეფიციარი-ი თესლი**; ჩემი აზრით საჭიროა გენმოდულირებული თესლების და კულტურების შემოტანის აკრძალვა; გაცილებით უფრო დიდი საფრთხის შემცველია გენმოდულირებული თესლი; **ბენეფიციარი-ი ორგანიზმი**; ცოცხალი გენმოდულირებული ორგანიზმები არის აგრესიული ჯიშის, რომელიც ხშირად ცხოველურ გენებსაც შეიცავს; სამედიცინო სფეროში გენმოდულირებული ორგანიზმები გამოიყენება წამლებისა და ვაქცინების წარმოებაში; კატეგორიულად უარყო მის პროდუქტებში გენმოდულირებული ორგანიზმების შემცველობა; **გენმოდულირებული პროდუქტი**; მოსწავლეებმა კომპლექსურად მიმოიხილეს გენმოდულირებული პროდუქციის წარმოებასთან, რეალიზაციასა და მათი მოხმარებით გამოწვეულ მავნე შედეგებთან დაკავშირებული საკითხები; მან ყურადღება გაამახვილა გენმოდულირებული

პროდუქტებით გაჭრობასთან ბრძოლის ევროპულ გამოცდილებაზე გენმოდლიფიცირებული პროდუქტების გამოყენებას თავისი დადებითი და უარყოფითი მხარეები აქვს; **ბენმოდიფიცირებადი პროდუქტია**; რამდენადაც ცნობილია, გენმოდლიფიცირებული პროდუქტია შედარებით იაფფასიანი; გენმოდლიფიცირებულ პროდუქტს საქართველოში, ევროპისგან განსხვავებით კანონმდებლობა არ არეგულირებს; **ბენმოდიფიცირებადი საკვები**; გენმოდლიფიცირებული საკვები დროთა განმავლობაში მუტაციას განიცდის, რაც მრავალი დაავადების გამომწვევი შეიძლება გახდეს; გენმოდლიფიცირებული საკვები მიიღება ხელოვნურად შექმნილი თესლებისგან; **ბენმოდიფიცირებადი ჯიშები**; გამოკითხული თბილისელების 80.2% საქართველოში გენმოდლიფიცირებული ჯიშების დანერგვის წინააღმდეგია; საქართველოში ინერგება არა მარტო ჰიბრიდული, არამედ გენმოდლიფიცირებული ჯიშებიც; საყურადღებოა, რომ ევროპას მკაცრი კრიტერიუმები აქვს გენმოდლიფიცირებული ჯიშების შემცველი კვების პროდუქტების მიმართ.

ქართული ენის განმარტებით ლექსიკონში სალექსიკონო სიტყვა-სტატია წარმოდგენილია **გენმოდლიფიცირებული** წარმოდგენილია ერთი მნიშვნელობით - **გენური ინჟინერიის შედეგი, გენშეცვლილი** და მინიჭებული აქვს სტილისუტი კვალიფიკაცია **ნეოლოგიზმი**. რაც შეეხება ქართული ენის ვებკორპუსის მიხედვით შედგენილ სიტყვა-სტატიას, გამოვყავით ამ სიტყვის მეორე მნიშვნელობა: **მდარე, უხარისხო; ახალ ამქალაქში მოგვევლინა თავისი მორიგი გენმოდლიფიცირებული ლექსით; ეს ებადაღებული გენმოდლიფიცირებული პოეზია ვის არ მოსწონს; ზედმეტად მორცხვი, გოიმურად კეკლუცი და საშუალოზე ლამაზი, გენმოდლიფიცირებული ადამიანური თვისებების საცავს წარმოადგენდა.** როგორც ჩანს, აღნიშნული მნიშვნელობა ამ სიტყვას, ახლანხანს გაუჩნდა, რაზეც ქართული ენის ვებკორპუსიდან ამოღებული საილუსტრაციო მასალა მოწმობს. გარდა ამ მნიშვნელობისა, ქართული ენის ვებკორპუსის მიხედვით შედგენილ სიტყვა-სტატიაში ასახული ისეთი შესიტყვებები, რომლებიც მოცემული არაა ქართული ენის განმარტებით ლექსიკონში, მაგალითად: **გენმოდლიფიცირებული თესლი; გაცილებით უფრო დიდი საფრთხის შემცველია გენმოდლიფიცირებული თესლი; გენმოდლიფიცირებული ორგანიზმი; სამედიცინო სფეროში გენმოდლიფიცირებული ორგანიზმები გამოიყენება წამლებისა და ვაქცინების წარმოებაში; გენმოდლიფიცირებული პროდუქტი; გენმოდლიფიცირებული პროდუქტების გამოყენებას თავისი დადებითი და უარყოფითი მხარეები აქვს;**

გენმოდიფიცირებული პროდუქცია; რამდენადაც ცნობილია,  
 გენმოდიფიცირებული პროდუქცია შედარებით იაფფასიანია;  
 გენმოდიფიცირებული საკვები; გენმოდიფიცირებული საკვები დროთა  
 განმავლობაში მუტაციას განიცდის, რაც მრავალი დაავადების გამომწვევი  
 შეიძლება გახდეს; გენმოდიფიცირებული ჯიში; გამოკითხული თბილისელების  
 80.2% საქართველოში გენმოდიფიცირებული ჯიშების დანერგვის წინააღმდეგია.  
 რაც შეეხება „ქართულ ლექსიკონს“, აღნიშნული სიტყვა არაა მოცემული.

## 22. სასწრაფო

### ქართული ენის განმარტებითი ლექსიკონი

**სასწრაფო (სასწრაფოსი) 1. სწრაფად შესასრულებელი, – საჩქარო, საშური.** *«გუშინ ვანო ქალაქიდან არ მოვიდა. ალბათ, სასწრაფო საქმე თუ გაუჩნდა»* (ს. შანშ.). *«იქ, ზევით, მიიმედ დაჭრილია! სასწრაფო შეველა სჭირია!»* (დ. სულ.). 2. **რაც ჩვეულებრივზე სწრაფად, დაუყოვნებლივ მიდის ან ხდება, – სწრაფად მოქმედი.** *«ეს სასწრაფო მატარებელია, უეჭველია, აქ ვინმე დიდი მოხელეც მიემგზავრება»* (ნ. ლორთქ.). *«თბილისიდან დეპეშაა, სასწრაფო»* (რ. გვეტ.). 3. **ძვ. თოფის წამლის ჭურჭელი (რქისა, ხისა და მისთ.), რომელსაც პირზე გაკეთებული აქვს საწყაო ერთი სროლა თოფის წამლისათვის.** *[ოთარაანთ ქვრივის] სახლში რომ შესულიყავით, ... ნახავდით ბუდით ჩამოკიდებულ კარგად მოზრდილ თოფს, ხმაღს, ... ხანჯალს, დამბახას, სასწრაფოს და საპირისწამლეს (ილია). || მასრების ჩანთა. [მთიულს] სასწრაფო და საპირისწამლე... მხარიდლივ პქონდა გადაკიდებული (ა. ყაზბ.). [ბიძა-ჩემმა] ამოაძრო სასწრაფოდან მასრები, სინჯავდა, წამალი იყო თუ არა შივა (ვაჟა).* 4. **კუთხ. (იმერ.) იგივეა, რაც სოლინარი. სასწრაფო არის ერთგვარი საკრავი. ის არის შემდგარი წვრილ-წვრილი ლერწმის მასრებისაგან, რომლებიც ერთად ბლის ქერქით არიან შეწნულნი: ამ საკრავს ნაზი და მრავალგვარი ხმა აქვს (გ. წერეთ.). სასწრაფოდ ზმნს. ჩქარა, მალე, დაუყოვნებლივ, სწრაფად.** *«კაცი გვიკვდება და სასწრაფოდ მოგვეშველე-თქო»* (ა. ბელ.). *ცეცო დატრიალდა, ... სასწრაფოდ დააგზნო... ცეცხლი (ს. მგალობ.). თუშმა... სასწრაფოდ შემოირტყა ხმალი (თ. რაზიკ.).*



## „ქართული ლექსიკონი“

**სასწრაფო** სასწრაფოს ძალიან სწრაფად შესასრულებელი, გადაუდებელი. სასწრაფო შეტყობინება. სასწრაფო შეკვეთა. სასწრაფო სამედიცინო დახმარება. სასწრაფოდ. ზმნიზ. ძალიან სწრაფად, დაუყოვნებლად. სასწრაფოდ უნდა დაბრუნდეს. სასწრაფოდ იყიდება ბინა.

## ქართული ენის ვებკორპუსი

სიხშირული ინდექსი: 7262

**სასწრაფო (სასწრაფოსი)** – არსებითი სახელი; იგივეა რაც სასწრაფო დახმარების მანქანა; ძლივს ლაპარაკობდა, ჰქონდა მაღალი სიციხე და ამის გამო გამოიძახა სასწრაფო; დაუყოვნებლივ უნდა მიმართოთ საავადმყოფოს ან გამოიძახოთ სასწრაფო; დედას ეხუმრობდნენ- მარი, ნულარ ჭამ ამდენს, სასწრაფო არ გაგვიხდეს გამოსადახებელი.

**სასწრაფო (სასწრაფოსი)** – ზედსართავი სახელი; 1. **სწრაფად შესასრულებელი**, – საჩქარო, საშური; თქვენს მიერ აღწერილი სიტუაცია სასწრაფო არ არის; დროულად დაფარავს გარდერობის სასწრაფო განახლების გაუთვალისწინებელ ხარჯებს; პრეზიდენტობის მოვალეობის შემსრულებელი ძალოვანი სტრუქტურების მეთაურებს სასწრაფო თათბირზე იწვევს; მაინც არაფერი გამოდიოდა, ვინაიდან რაღაც-რაღაცები სასწრაფო იყო და დრო არ ითმენდა; 2. **რაც ჩვეულებრივზე სწრაფად, დაუყოვნებლივ მიდის ან ხდება**, – **სწრაფად მოქმედი**; ერწოს ზონაზე რომ ნახოთ, რა მდგომარეობაში სასწრაფო დახმარების პუნქტი გული გაგისკდებათ; დაუკავშირდა სასწრაფო დახმარების ოპერატიული სამსახურის უფროსს; **სასწრაფო დახმარება**; გულის მწვავე უკმარისობის ძირითადი ნიშნების გამოვლენისთანავე უნდა გამოვიძახოთ სასწრაფო დახმარება; ბლოგზე ჩვენ განვათავსებთ ყველა უახლოეს ინფორმაციას თუ ვის სჭირდება სასწრაფო დახმარება; მსგავსი სიმპტომების არსებობისას უნდა მივმართოთ სასწრაფო დახმარებას ან სტაციონარს; **სასწრაფო დახმარების ბრიგადა**; დროულად გამოვიძახოთ სასწრაფო დახმარების ბრიგადა; სასწრაფო დახმარების ბრიგადა გზაშია; სასწრაფო დახმარების ბრიგადამ, სამი დაშავებული საუნივერსიტეტო კლინიკასა და პირველ კლინიკურ საავადმყოფოში გადაიყვანა; **სასწრაფო დახმარების ექიმი**; სასწრაფო დახმარების ექიმებმა მისი გადარჩენა ვერ შეძლეს; სასწრაფო

დახმარების ექიმი ასევე ვაღდებულება, პაციენტს ტკივილგამაყუჩებელი პრეპარატი გაუკეთოს; სასწრაფო დახმარების ექიმებს დიდი გამოცდილება აქვთ; **სასწრაფო დახმარების მანქანა;** მიმდებარე ტერიტორიაზე იმორიგეებს სასწრაფო დახმარების მანქანა; ადვილზე მისული სასწრაფო დახმარების მანქანა აუცილებლად რეკავს საპარტულო სამსახურში; საცხოვრებელი კორპუსების ახლოს, სასწრაფო დახმარების მანქანა დგას; სასწრაფო დახმარების მანქანა მაშინ მოდის, როცა ის ავადმყოფს აღარ ჭირდება; **სასწრაფო ოპერაცია;** ერთხელ ჩემს მეუღლეს სასწრაფო ოპერაცია დასჭირდა; მას გადაუდებელი დახმარების ცენტრში სასწრაფო ოპერაცია ჩაუტარდა; ბიუჯეტის ახალ პროექტში გაზრდილია გეგმიური და სასწრაფო ოპერაციების დაფინანსება; **სასწრაფო სამედიცინო დახმარება;** სამაშველო სამსახურის და სასწრაფო სამედიცინო დახმარების ბრიგადები; მძიმე შემთხვევაში აუცილებელია სასწრაფო სამედიცინო დახმარების გამოძახება; ჯანდაცვის პროგრამებს გამოეყო სასწრაფო სამედიცინო დახმარების პროგრამისთვის გათვალისწინებული თანხები; **სასწრაფო სამედიცინო ჩარევა;** ფილტვების შეშუპება სასწრაფო სამედიცინო ჩარევას მოითხოვს; ცხვირის ძვიდის პემეტომაც სასწრაფო სამედიცინო ჩარევას საჭიროებს; ანაფილაქსიური შოკი სასწრაფო სამედიცინო ჩარევას მოითხოვს; **სასწრაფო ჰოსპიტალიზაცია;** თვითნებითი აბორტის დროს აუცილებელია სასწრაფო ჰოსპიტალიზაცია; საჭიროა სასწრაფო ჰოსპიტალიზაცია და ადეკვატური ზომების მიღება საგადალო შედეგების თავიდან ასაცილებლად; მუცლის ღრუში ტკივილის არსებობისას აუცილებელია სასწრაფო ჰოსპიტალიზაცია ქირურგიულ განყოფილებაში.

ქართული ენის განმარტებით ლექსიკონში სალექსიკონო სიტყვა-სტატია **სასწრაფო** წარმოდგენილია ოთხი მნიშვნელობით, მათ შორის დიალექტური მასალიდან. ხოლო „ქართულ ლექსიკონში“ ამ სიტყვის ერთი მნიშვნელობაა მოცემული. ქართული ენის ვებკორპუსის მიხედვით შედგენილ სიტყვა-სტატიაში, ცალკე სიტყვა-სტატიად გამოვიტანეთ მოცემული სიტყვა როგორც არსებითი სახელი, შემდეგი მნიშვნელობით: **იგივეა რაც სასწრაფო დახმარების მანქანა;** ძლივს ლაპარაკობდა, ჰქონდა მაღალი სიცხე და ამის გამო გამოიძახა სასწრაფო; დაუყოვნებლივ უნდა მიმართოთ საავადმყოფოს ან გამოიძახოთ სასწრაფო; დედას ეხუმრებოდნენ- მარი, ნულარ ჭამ ამდენს, სასწრაფო არ

გაგვიხდეს გამოსაძახებელიო. აღსანიშნავია, რომ ზედსართავ სახელში, ქართული ენის ვებკორპუსის მიხედვით შედგენილ სიტყვა-სტატიაში ასახვა ვერ პოვა დიალექტურმა მასალამ ქართული ენის განმარტებითი ლექსიკონისაგან განსხვავებით, თუმცა მასში წარმოდგენილია ისეთი შესიტყვებები, რომლებიც თანამედროვე ქართულ ენაში აქტიურად გამოიყენება და მოცემული არ არის ქართული ენის განმარტებით ლექსიკონში. ესენია: **სასწრაფო დახმარება**; *მსგავსი სიმპტომების არსებობისას უნდა მივმართოთ სასწრაფო დახმარებას ან სტაციონარს*; **სასწრაფო დახმარების ბრიგადა**; *სასწრაფო დახმარების ბრიგადა ვაშია*; **სასწრაფო დახმარების ექიმი**; *სასწრაფო დახმარების ექიმებმა მისი გადარჩენა ვერ შეძლეს*; **სასწრაფო დახმარების მანქანა**; *მიმდებარე ტერიტორიაზე იმორიგევენ სასწრაფო დახმარების მანქანა*; **სასწრაფო ოპერაცია**; *ბიუჯეტის ახალ პროექტში გაზრდილია გეგმიური და სასწრაფო ოპერაციების დაფინანსება*; **სასწრაფო სამედიცინო დახმარება**; *მიიმე შემთხვევაში აუცილებელია სასწრაფო სამედიცინო დახმარების გამოძახება*; **სასწრაფო სამედიცინო ჩარევა**; *ფილტვების შეშუპება სასწრაფო სამედიცინო ჩარევას მოითხოვს*; **სასწრაფო ჰოსპიტალიზაცია**; *თვითნებითი აბორტის დროს აუცილებელია სასწრაფო ჰოსპიტალიზაცია*.

## 23. ტენდერი

### ქართული ენის განმარტებითი ლექსიკონი

**ტენდერ-ი** (ტენდერისა) [ინგ. tender] *სპეც. ორთქლმავლის ნაწილი ან საგანგებოდ მოწყობილი მისაბმელი ვაგონი, რომელშიც ორთქლმავლისათვის საჭირო წყლის, სათბობისა და სხვა მასალის მარაგი თავსდება. ზეინკლებს ევალება ტენდერის ახსნა, აწევა («კომ.»).*

### „ქართული ლექსიკონი“

**ტენდერი** *ტენდერის* ღია კონკურსი რაიმეს შესყიდვაზე ან რომელიმე საქმის შესრულებაზე უფლების მოსაპოვებლად. *ტენდერი გამოცხადდა. ტენდერში გამარჯვებული კომპანია.*

## ქართული ენის ვებკორპუსი

სიხშირული ინდექსი: 1023

**ტენდერ-0 (ტენდერისა) – სახელწმიფო თუ არასახელმწიფო** ორგანიზაციის მიერ გამოცხადებული ღია კონკურსი ამა თუ იმ ტექნიკური მოწყობილობისა და მომსახურების შესყიდვის თაობაზე; *დედაქალაქის მერიამ ბავშთა ქალაქ „მზიურის“ რებილიტაციაზე ტენდერი 16 მაისს გამოაცხადა; წყალტუბოს მუნიციპალიტეტის მიერ გამოცხადდა გამარტივებული ელექტრონული ტენდერი 2 ერთეულ ნაგავმზიდი ავტომანქანის შესყიდვაზე; გამგეობას უფლება მისცეს კულტურული ღონისძიების მომსახურებისთვის ტენდერი გამარტივებული წესით ჩაატაროს; „მაგთიკომმა“ ცხელი ხაზის მომსახურებაზე თბილისის სატრანსპორტო კომპანიის ტენდერი მოივო.*

ქართული ენის განმარტებით ლექსიკონში აღნიშნული სიტყვის მხოლოდ ერთი მნიშვნელობაა მოცემული, რაც ორთქმავლის ნაწილს ან საგანგებობდ მისაბმელ ვაგონს უკავშირდება. ქართული ენის ვებკორპუსის მიერ შედგენილ სიტყვა-სტატიაში კი პირიქით, ასახვა არ პოვა ამ მნიშვნელობამ, მაგრამ მასში მოცემული ის მნიშვნელობა, რაც ასე აქტიურდ გამოიყენება დღეს: **სახელწმიფო თუ არასახელმწიფო ორგანიზაციის მიერ გამოცხადებული კონკურსი ამა თუ იმ ტექნიკური მოწყობილობისა და მომსახურების შესყიდვის თაობაზე; დედაქალაქის მერიამ ბავშთა ქალაქ „მზიურის“ რებილიტაციაზე ტენდერი 16 მაისს გამოაცხადა.** აღსანიშნავია, რომ „ქართულ ლექსიკონში“ მოცემულია ამ სიტყვის ის მნიშვნელობა, რაც ვებკორპუსის მიხედვით შედგენილ სიტყვა-სტატიაშია წარმოდგენილი.

## 24. ბალანსი

### ქართული ენის განმარტებითი ლექსიკონი

**ბალანსი (ბალანსისა)** სახელი გადანერგავს, გადაინერგება ზმნათა მოქმედებისა.

#### „ქართული ლექსიკონი“

**ბალანსი** *გადანერგვის* (გადანერგავს, გადაუნერგავს, გადაინერგავს, გადაინერგება) ცოცხალი ორგანიზმის რომელიმე ორგანოს, ქსოვილის და მისთ. ამოკვეთა და სხვა ორგანიზმში დაზიანებული ორგანოს ნაცვლად გადატანა მისი სასიცოცხლო ფუნქციის აღდგენის მიზნით. *თმის გადანერგვა. გული გადაუნერგეს. გადანერგილი. რაც გადანერგეს. გადანერგილი კანი.*

#### ქართული ენის ვებკორპუსი

სიხშირული ინდექსი: 1022

**ბალანსი (ბალანსისა)** ამა თუ იმ ორგანოს, უჯრედის, ქსოვილის გადატანა; რეციდივის შემთხვევაში მკურნალობის იდეალური ვარიანტია ძვლის ტვინის გადანერგვა; სადღეისოდ მკურნალობის ოქროს სტანდარტად მიჩნეულია ღვიძლის გადანერგვა; ზოგჯერ ამ სტადიაშიც ტარდება ძვლის ტვინის გადანერგვა; ძლიერი დამწვრობის შემთხვევაში შესაძლოა დაგჭირდეთ კანის გადანერგვა; ამის შემდეგ ხდება ჯანმრთელი ღეროვანი უჯრედების გადანერგვა; თმის გადანერგვის ეფექტურობა მან თავის თავზე გამოსცადა.

ქართული ენის განმარტებით ლექსიკონსა და „ქართულ ლექსიკონში“ აღნიშნული სიტყვის ერთი მნიშვნელობაა მოცემული, თუმცა მოცემული არაა ვრცელი საილუსტრაციო მასალა. ქართული ენის ვებკორპუსის მიხედვით შედგენილ სიტყვა-სტატიაში კი მოცემული საილუსტრაციო მასალა სხადასხვა დარგიდან: რეციდივის შემთხვევაში მკურნალობის იდეალური ვარიანტია ძვლის ტვინის გადანერგვა; სადღეისოდ მკურნალობის ოქროს სტანდარტად მიჩნეულია ღვიძლის გადანერგვა; ზოგჯერ ამ სტადიაშიც ტარდება ძვლის ტვინის გადანერგვა; ძლიერი დამწვრობის შემთხვევაში შესაძლოა დაგჭირდეთ კანის გადანერგვა; ამის შემდეგ ხდება ჯანმრთელი ღეროვანი უჯრედების გადანერგვა; თმის გადანერგვის ეფექტურობა მან თავის თავზე გამოსცადა.

## 25. სცენარი

### ქართული ენის განმარტებითი ლექსიკონი

სცენარ-0 (სცენარისა) [იტალ. scenario] 1. დრამატული ნაწარმოები, რომლის მიხედვითაც იღებება კინოსურათი. 2. თეატრ. დრამატული ნაწარმოების სიუჟეტური სქემა, მონახაზი, რომელსაც მსახიობები სპექტაკლის მსვლელობის დროს იმპროვიზაციულად ავსებენ. 3. პიესის მოქმედ პირთა სია, სადაც ნაჩვენებია თითოეული მათგანის სცენაზე გამოსვლის დრო და რიგი.

### „ქართული ლექსიკონი“

სცენარი სცენარის მხატვრული ნაწარმოები, რომლის მიხედვითაც იღებენ ფილმს ან დგამენ წარმოდგენებს. სცენარის ავტორი.

### ქართული ენის ვებკორპუსი

სიხშირული ინდექსი: 1020

სცენარ-ი (სცენარისა) 1. მხატვრული ნაწარმოების სიუჟეტური სქემა, მონახაზი, რომელსაც მსახიობების ფილმის, სპექტაკლის დროს ავსებენ; ერთადერთი გამოსავალი ის დარჩება, რომ მე თვითონ დავწერო სცენარი, გადავიღო ფილმი და თავადვე ვითამაშო; თურმე, მეუღლეს რამდენიმე კინორეჟისორთან მიუტანია სცენარი, და ყველამ, არც მეტი, არც ნაკლები, უარით გაისტუმრა; პიტერმა სცენარი ჰოლივუდის სხვა კომპანიებს მიუტანა; 2. ამა თუ იმ მოვლენათა განვითარების სურათი; აი, სცენარი, არჩევნები გაყალბდება, ხალხი ქუჩაში გამოვა; ძნელია, წარმოიდგინო იმაზე უარესი სცენარი, ვიდრე თვითმფრინავის მარცხი და შუა ოკეანეში ჩავარდნა; ახლო აღმოსავლეთში უკვე გათამაშდა ის სცენარი, რაც მთელ მსოფლიოშია მოსალოდნელია; ქართულმა საზოგადოებამ თავად შეამზადა ეთნიკური კონფლიქტების ის სცენარი, რომელშიც საქართველო სრულიად ლოგიკურად აღმოჩნდა.

როგორც ვხედავთ, ქართული ენის განმარტებით ლექსიკონში მოცემული სიტყვის სამი მნიშვნელობაა გამოტანილი, თუმცა არ არის საილუსტრაციო მასალა განსხვავებით ქართული ენის ვებკორპუსის მიხედვით შედგენილ სიტყვა-სტატიისაგან. ხოლო „ქართულ ლექსიკონში“ მოცემული სიტყვის ერთი მნიშვნელობაა გამოტანილი, ქართული ენის ვებკორპუსში კი ამ სიტყვის ორი

მნიშვნელობა, მათ შორის ისეთი მნიშვნელობა, რომელიც ასახული არ არის არც ქართული ენის განმარტებით ლექსიკონში და არც „ქართულ ლექსიკონში“. ეს მნიშვნელობაა: **ამა თუ იმ მოვლენათა განვითარების სურათი; აი, სცენარი, არჩევნები გაყალბდება, ხალხი ქუჩაში გამოვა; ძნელია, წარმოიდგინო იმაზე უარესი სცენარი, ვიდრე თვითმფრინავის მარცხი და შუა ოკეანეში ჩავარდნა.**

## 26. გენეტიკა

### ქართული ენის განმარტებითი ლექსიკონი

**გენეტიკა (გენეტიკოსა)** [ბერძ. genetikos წარმოშობასთან დაკავშირებული] ბიოლოგიის დარგი, რომელიც შეისწავლის ორგანიზმის განვითარებასა და მემკვიდრეობითობას. *სიმსუქნის ერთ-ერთი ფაქტორი ადამიანის ინდივიდუალური გენეტიკაა («კვ. ქრონი»).* მაინტერესებს მხოლოდ ის, რაც გენეტიკის ზოგადთეორიულ საკითხებს ეხება (თ. ბუაჩ).

### „ქართული ლექსიკონი“

**გენეტიკა** გენეტიკის მეცნიერება, რომელიც შეისწავლის ორგანიზმების მემკვიდრეობისა და ცვალებადობის კანონებს. *გენეტიკის კურსი.*

### ქართული ენის ვებკორპუსი

სიხშირული ინდექსი: 839

**გენეტიკა (გენეტიკოსა)** 1. **გენეტიკური კოდი, ინფორმაცია, მემკვიდრეობითობა;** *ისინი დამეხმარენ იმით, რომ გამაჩინეს და მათი გენეტიკა გამოძევა; თუმცა ადამიანის ხასიათს გენეტიკა განაპირობებს და პლასტიკური ქირურგიით ამ ხასიათს ვერ გადააკეთებ; დოქტორმა ნორმან ბორლაუგმა, შეისწავლა მცენარეების გენეტიკა და მიღებულ ინფორმაციაზე დაყრდნობით შექმნა ახალი თაობის ხორბალი; არც მთლად მასეა საქმე, როცა ჩემი გურულსვანური გენეტიკა თავს მახსენებს;* 2. **ბიოლოგიის დარგი, რომელიც შეისწავლის ორგანიზმის განვითარებასა და მემკვიდრეობითობას;** *დაეუფლები ისეთ ფუნდამენტურ და გამოყენებით ბიოლოგიურ დისციპლინებს, როგორებიცაა: გენეტიკა, მოლეკულური ბიოლოგია; უნივერსიტეტში 2008 წლიდან კითხულობს ლექციებს სამედიცინო გენეტიკა, ონკოლოგია;*

ციტოლოგია და გენეტიკა, ბიოლოგიის ეს ორი დარგი, მჭიდროდ დაუკავშირდა ერთმანეთს, **მოლეკულური გენეტიკა**; კვლევას ახორციელებს აგრარული უნივერსიტეტის ფარგლებში არსებული მოლეკულური გენეტიკის ინსტიტუტი; ქალბატონმა იამ თავიდან მოკლედ მიმოიხილა მოლეკულური გენეტიკა.

ქართული ენის განმარტებით ლექსიკონსა და „ქართულ ლექსიკონში“ აღნიშნული სიტყვა-სტატია მოცემულია მხოლოდ ერთ მნიშვნელობით, ხოლო ქართული ენის ვებკორპუსის მიხედვით შედგენილი სიტყვა-სტატიაში გამოიყოფა ორი მნიშვნელობა: 1) **გენეტიკური კოდი, ინფორმაცია, მემკვიდრეობითობა**; თუმცა ადამიანის ხასიათს გენეტიკა განაპირობებს და პლასტიკური ქირურგიით ამ ხასიათს ვერ გადააკეთებ და 2) **ბიოლოგიის დარგი, რომელიც შეისწავლის ორგანიზმის განვითარებასა და მემკვიდრეობითობას**; უნივერსიტეტში 2008 წლიდან კითხულობს ლექციებს სამედიცინო გენეტიკა, ონკოლოგია. გარდა ამისა, გამოტანილია შესიტყვება, **მოლეკულური გენეტიკა**; კვლევას ახორციელებს აგრარული უნივერსიტეტის ფარგლებში არსებული მოლეკულური გენეტიკის ინსტიტუტი.

## 27. ბენეტიკური

### ქართული ენის განმარტებითი ლექსიკონი

**ბენეტიკური-ი (გენეტიკურისა) გენეტიკასთან დაკავშირებული, რაიმეს წარმოშობის, განვითარების შემსწავლელი. გენეტიკური მინერალოგია.** – აქ ბრმა მიბაძვაზე, რა თქმა უნდა, ლაპარაკიც არ შეიძლება, ესაა მხოლოდ გენეტიკური ნათესაობა («ლიტ. საქ»). ინსტიტუტი უზნაძის სიცოცხლეში სამი – ზოგადი ფსიქოლოგიის, გენეტიკური ფსიქოლოგიისა და პათოფსიქოლოგიის განყოფილებებისაგან შედგებოდა (რ. ნათ.). ქართველ კაცში სწრაფად იჩინა თავი თავისუფლების გენეტიკურმა წყურვილმა (დ. მჭედლ.). **გენეტიკურად ზმნს.** წარმოშობის მხრივ, წარმოშობის თვალსაზრისით. ვფიქრობ, [ეს] გენეტიკურად გაპირობებული რაიმე აუხსნელი თვისება უნდა ყოფილიყო (თ. ბუაჩ.). **გენეტიკური კოდი** შთამომავლობითი «ინფორმაცია», რომელიც გენებით გადაეცემა. ბებიაჩემის ოპტიმიზმი გენეტიკური კოდით გადმომეცა (ჭ. ამირეჯ.). ქართული სული, ეროვნული ტრადიციები ჩვენს ხალხშია ფესვადგმული და მამებიდან შვილებს გენეტიკური კოდით გადაეცემათ («ახ.-დას.»). **გენეტიკური**



ფონდი მოცემული პოპულაციის, პოპულაციათა ჯგუფის ან სახეობის გენების ერთობლიობა. მწამს, რომ ადამიანის სული არის უკვდავი – იგი გენეტიკური ფონდია, შთამომავლობით გადადის და არამარტო შთამომავლობით (ჭ. ამირეჯ.).

### „ქართული ლექსიკონი“

**ბენეტიკური გენეტიკურის 1. გენეტიკასთან დაკავშირებული. გენეტიკური კვლევა. 2. გენებთან დაკავშირებული. გენეტიკური დაავადება.**

### ქართული ენის ვებკორპუსი

სიხშირული ინდექსი: 987

**ბენეტიკურ-ი (ბენეტიკურისა) გენეტიკასთან დაკავშირებული, რაიმეს წარმოშობის, განვითარების შემსწავლელი; სარკოდოზს საფუძვლად უდევს გენეტიკური, ეკოლოგიური, ინფექციური და იმუნოლოგიური ფაქტორები; ამ პროცესის შეფერხება შესაძლოა გენეტიკური ან ჰორმონული დარღვევებით იყოს განპირობებული; ევოლუციის მანძილზე, ორგანიზმში გენეტიკური ცვლილებები ხდება; გენეტიკური ალგორითმი; გენეტიკური ალგორითმების მეთოდოლოგიის არსი ევოლუციის ბუნებრივი შერჩევის პრინციპს ეფუძნება; გენეტიკური ანალიზი; გენეტიკური ანალიზი აჩვენებს რომ ტუტანხამონი არაერთი სნეულებით იყო დაავადებული; მსოფლიოს განვითარებულ ქვეყნებში პაციენტებს გენეტიკური ანალიზი უტარდებათ; გენეტიკური ანალიზით დადგენილია, რომ დაავადების გამომწვევი გენი ლოკალიზებულია მე-13 ქრომოსომაზე; გენეტიკური ანომალია; ასეთი გენეტიკური ანომალიით 20 000 ახალშობილი გაუიდან ერთი იბადება; მრავალი გენეტიკური ანომალიისას, რომელსაც ამენორეა ახასიათებს, დარღვეულია შიგნითა სასქესო ორგანოების განვითარება; არცთუ იშვიათად ლეიკოპენია იმუნიტეტის გენეტიკური ანომალიით არის განპირობებული; გენეტიკური აპარატი; გენეტიკური აპარატის - დნმ-ს დაზიანება ვლინდება მხოლოდ გარემოს არახელსაყრელი ზემოქმედების შემდეგ; უჯრედის გენეტიკური აპარატის ანუ დნმ-ს იმგვარი ცვლილება, რომელიც უჯრედს უკონტროლოდ გამრავლებას აიძულებს; გენეტიკურია დაავადება, რომელსაც ადამიანის გენეტიკური აპარატის ანუ დნმ-ს მოლეკულის დაზიანება იწვევს; გენეტიკური განწყობა; ბავშვები მზა პიროვნებებად არ იბადებიან, მათ ჩამოყალიბებაში თავიანთ როლს ასრულებენ როგორც**

გენეტიკური განწყობა, ასევე გარემო; შაქრიანი დიაბეტის განვითარებისთვის რისკის ფაქტორებია გენეტიკური განწყობა, პათოლოგიური ორსულობა; მნიშვნელოვან როლს განსაზღვრავს დაავადების განვითარებაში გენეტიკური განწყობა; **გენეტიკური დაავადება**; პაციენტს ჰქონდა იშვიათი გენეტიკური დაავადება, ფანკონის ანემია; გენეტიკური დაავადება შესაძლოა ოჯახის მხოლოდ ერთ წევრს გამოაჩნდეს; გენეტიკური დაავადებები სამ ჯგუფად იყოფა, იმის მიხედვით, რომელი ფაქტორი ახდენს მათ ჩამოყალიბებასა და განვითარებაზე უფრო მეტ გავლენას - მემკვიდრეობითი თუ გარეგანი; **გენეტიკური დარღვევა**; დღეისთვის დაავადების სავარაუდო მიზეზად გენეტიკური დარღვევა მოიაზრება; ლეშ-ნიჰანის სინდრომი გენეტიკური დარღვევაა, რომელიც 380 000-დან ერთ შემთხვევაში იხენს; ეს მუტაცია ერთგვარი გენეტიკური დარღვევაა, რომელიც უჯრედებს შიდსის მიმართ ბრძოლისუნარიანს ხდის; **გენეტიკური დეფექტი**; სამედიცინო მიზეზად ითვლება შემთხვევები, როდესაც ნაყოფს გენეტიკური დეფექტი აქვს; სადღეისოდ აღწერილია ალფა-1-ანტიტრიფსინის გენის 100-მდე გენეტიკური დეფექტი; პირველი ჰიპოლაქტაზია გენეტიკური დეფექტით არის განპირობებული; **გენეტიკური ვარიაცია**; სქესობრივი გამრავლება ზრდის გენეტიკურ ვარიაციას და ქმნის იმ მრავალფეროვნებას, რომელსაც აკონტროლებს ბუნებრივი გადარჩევა; დღესდღეობით ცნობილია, რომ მათი უნიკალურობის მიზეზი ჩვენს სახეობაში არსებული გენეტიკური ვარიაციებია; ცხენ-ირმის შემთხვევაში შესაძლოა ნაშიერი არაბუნებრივი გენეტიკური ვარიაციებით გამოირჩეოდეს; **გენეტიკური ინჟინერია**; შემოიტანეს გენეტიკური ინჟინერიით მიღებული კარტოფილის სამი ჯიში; ბიომეურნეობაში კატეგორიულად აკრძალულია გენეტიკური ინჟინერიის ტექნოლოგიით მიღებული ორგანიზმების გამოყენება; გენეტიკური ინჟინერიაც მოწოდების სიმადლეზე იყო; **გენეტიკური ინფორმაცია**; ადამიანის გენეტიკური ინფორმაცია ინახება მოლეკულაში, რომელსაც დნმ ეწოდება; ვირუსს დიდი წნევით შეჰყავს თავისი გენეტიკური ინფორმაცია მასპინძელ უჯრედში, რომელიც შემდეგ პარაზიტს ამრავლებს და ინფექციის გავრცელებას ხელს უწყობს; ახალი გენეტიკური ინფორმაცია მასში გენების ჰორიზონტალური გადაცემით შედის; **გენეტიკური კოდი**; უკვე გაშიფრულია ამ საშიში მიკრობის გენეტიკური კოდი; მატრიქსის აქტიურობას ყოველ ადამიანში მისთვის დამახასიათებელი გენეტიკური კოდი განსაზღვრავს; ჩვენი გენეტიკური კოდი ჯერ კიდევ შეიცავს ეგოისტურ და აგრესიულ ინსტინქტებს, რომლებიც

წარსულში გადარჩენისთვის იყო საჭირო; მისმა გამოკვლევებმა კაცობრიობას გენეტიკური კოდის აღმოჩენის საშუალება მისცა; **გენეტიკური მარკერი**; გამოვლინდა გენეტიკური მარკერები, რომლებიც ხშირად არის ასოცირებული არასპეციფიკურ წყლულოვან კოლიტთან; ამ სინდრომთა უმრავლესობისთვის განსაზღვრულია გენეტიკური მარკერები, მათი ფუნქცია ბოლომდე მაინც არ არის შესწავლილი; **გენეტიკური მეხსიერება**; ახლა ვმუშაობ გენეტიკური მეხსიერების თემაზე; ამბობენ ადამიანებს გენეტიკური მეხსიერებაც გააჩნიათო, არქეტიპებიც; როგორც ჩანს, გენეტიკური მეხსიერება, ძველ ნათესაურ კოდს მაინც ინარჩუნებს; **გენეტიკური მიდრეკილება**; სიმსუქნისადმი გენეტიკური მიდრეკილება ის ძირითადი ფაქტორებია, რომლებიც სიმსუქნეს უღვეს საფუძვლად; სიმსუქნისადმი გენეტიკური მიდრეკილება ის ძირითადი ფაქტორებია, რომლებიც სიმსუქნეს უღვეს საფუძვლად; გენეტიკური მიდრეკილებით გამოწვეული დარღვევების შედეგად მცირდება პლაცენტური სისხლის მიმოქცევა; **გენეტიკური მუტაცია**; ზოგიერთი გენეტიკური მუტაცია ზრდის ცელიაკიის ალბათობას; გენეტიკური მუტაცია მასივნალიზებელ ცილებს თავის ტვინის უჯრედების ერთმანეთთან დაკავშირებაში უშლის ხელს; ბიოლოგიური მრავალფეროვნება გამოიხატება გენეტიკურ მუტაციებში; **გენეტიკური პათოლოგია**; დაუნის სინდრომი გენეტიკური პათოლოგიაა; გვიან ასაკში მაღალია ახალშობილთა გენეტიკური პათოლოგიებით დაავადების ალბათობა; ბოლო ხანს საზოგადოდ გენეტიკური პათოლოგიების გახშირება შეიმჩნევა; **გენეტიკური რესურსი**; ვაზის ადგილობრივი გენეტიკური რესურსების იდენტიფიკაცია; აკად. ნ. ვავილოვის მიერ დაწყებული გენეტიკური რესურსების შეგროვება; ლექცია მოიცავდა ინფორმაციას ვაზის ბიომრავალფეროვნებისა და გენეტიკური რესურსების შესახებ; **გენეტიკური ტესტი**; გენეტიკური ტესტების პასუხები დადასტურებულია: ახალგაზრდა მამაკაცის მამა ეხნატონია; ახალ აღმოჩენებზე დაყრდნობით მათ შექმნეს გენეტიკური ტესტი; **გენეტიკური ტესტირება**; ამ მეთოდს დედის სისხლში მოცირკულირე ნაყოფის უჯრედების გენეტიკური ტესტირება ეწოდება; საჭიროა, უკანონოდ გამოცხადდეს ადამიანის ნებისმიერი უჯრედის გენეტიკური ტესტირება მფლობელის თანხმობის გარეშე; გენეტიკური ტესტირება ყოველწლიურად უნდა ჩატარდეს ოჯახის 6 წელს გადაცილებულ ყველა წევრს; **გენეტიკური ფაქტორი**; მხედველობის გაუარესებას მნიშვნელოვანწილად განსაზღვრავს გენეტიკური ფაქტორები; გენეტიკური ფაქტორები გავლენას ახდენს იმ პორმონებზე, რომლებიც ცხიმოვანი ცვლის

რეგულაციაში მონაწილეობს; კვლევის თანახმად, სიზარმაცეს შესაძლოა გენეტიკური ფაქტორები განაპირობებდეს; **გენეტიკური ფონი**; რტერიული წნევის ურთიერთკავშირზე მრავალი ფაქტორი ახდენს გავლენას, მათ შორის - ორგანიზმის თავისებურება, გენეტიკური ფონი; დატვირთული გენეტიკური ფონი მაქვს.

## **ქართული ენის ვებკორპუსი**

სიხშირული ინდექსი: 657

### **28. ბენეტიკურად**

**ბენეტიკურად** – ზმნ. წარმოშობის/ გენეტიკის მხრივ, თვალსაზრისით; თავმდაბლობა ჩვენში არ გამოდის, არა ვართ თავმდაბალი გენეტიკურად; შეშუპებული ქუთუთოებიც გენეტიკურად მაქვს დედაჩემის მხრიდან; სქესობრივი ლტოლვის მიმართულება, როგორც ცხოველებში, ასევე ადამიანებში გენეტიკურად არის განპირობებული; **გენეტიკურად მოდიფიცირებული**; გენეტიკურად მოდიფიცირებულ თავგებს ნორმაზე გაცილებით პატარა სათესლეები განუვითარდათ; მოხსენება შეეხება გენეტიკურად მოდიფიცირებულ ორგანიზმებს (გმო); დღეს მსოფლიოში ყოველწლიურად იზრდება გენეტიკურად მოდიფიცირებული პროდუქციის წარმოება.

ქართული ენის განმარტებით ლექსიკონო სალექსიკონო ერთეულისატვის **გენეტიკური** გამოტანილია ერთი მნიშვნელობა, ხოლო „ქართულ ლექსიკონში“ ამავე სიტყვის ორი მნიშვნელობა გვხვდება. ასევე ერთი მნიშვნელობაა გამოტანილი ქართული ენის ვებკორპუსის მიხედვით შედგენილ სიტყვა-სტატიაში, თუმცა აქ ყურადღებას იპყრობს საილუსტრაციო მასალის მრავალფეროვნება და ისეთი შესიტყვებები, რომლებიც არ არის მოცემული ქართული ენის განმარტებით ლექსიკონში. ესენია: **გენეტიკური ალგორითმი**; **გენეტიკური ალგორითმების მეთოდოლოგიის არსი ევოლუციის ბუნებრივი შერჩევის პრინციპს ეფუძნება**; **გენეტიკური ანალიზი**; **გენეტიკური ანალიზი აჩვენებს რომ ტუტანხამონი არაერთი სნეულებით იყო დაავადებული**; **გენეტიკური ანომალია**; **ასეთი გენეტიკური ანომალიით 20 000 ახალშობილი ვაჟიდან ერთი იბადება**; **გენეტიკური აპარატი**; **გენეტიკური აპარატის - დნმ-ს დაზიანება ვლინდება მხოლოდ გარემოს არახელსაყრელი ზემოქმედების შემდეგ**;

**გენეტიკური განწყობა;** მნიშვნელოვან როლს განსაზღვრავს დაავადების განვითარებაში გენეტიკური განწყობა; **გენეტიკური დაავადება;** პაციენტს ჰქონდა იშვიათი გენეტიკური დაავადება, ფანკონის ანემია; **გენეტიკური დარღვევა;** დღეისთვის დაავადების სავარაუდო მიზეზად გენეტიკური დარღვევა მოიაზრება; **გენეტიკური დეფექტი;** სადღეისოდ აღწერილია ალფა-1-ანტიტრიფსინის გენის 100-მდე გენეტიკური დეფექტი; **გენეტიკური ვარიაცია;** ცხენ-ირმის შემთხვევაში შესაძლოა ნაშიერი არაბუნებრივი გენეტიკური ვარიაციებით გამოირჩეოდეს; **გენეტიკური ინჟინერია;** ბიომეურნეობაში კატეგორიულად აკრძალულია გენეტიკური ინჟინერიის ტექნოლოგიით მიღებული ორგანიზმების გამოყენება; **გენეტიკური ინფორმაცია;** ადამიანის გენეტიკური ინფორმაცია ინახება მოლეკულაში, რომელსაც დნმ ეწოდება; **გენეტიკური კოდი;** უკვე გაშიფრულია ამ საშიში მიკრობის გენეტიკური კოდი; **გენეტიკური მარკერი;** გამოვლინდა გენეტიკური მარკერები, რომლებიც ხშირად არის ასოცირებული არასპეციფიკურ წყლულოვან კოლიტთან; **გენეტიკური მენსიერება;** ახლა ვმუშაობ გენეტიკური მენსიერების თემაზე; **გენეტიკური მიდრეკილება;** სიმსუქნისადმი გენეტიკური მიდრეკილება ის ძირითადი ფაქტორებია, რომლებიც სიმსუქნეს უღვევს საფუძვლად; **გენეტიკური მუტაცია;** ზოგიერთი გენეტიკური მუტაცია ზრდის ცელიაკიის ალბათობას; **გენეტიკური პათოლოგია;** დაუნის სინდრომი გენეტიკური პათოლოგიაა; **გენეტიკური რესურსი;** ვაზის ადგილობრივი გენეტიკური რესურსების იდენტიფიკაცია; **გენეტიკური ტესტი;** გენეტიკური ტესტების პასუხები დადასტურებულია: ახალგაზრდა მამაკაცის მამა ეხნატონია; **გენეტიკური ტესტირება;** ამ მეთოდს დედის სისხლში მოციროკულირე ნაყოფის უჯრედების გენეტიკური ტესტირება ეწოდება. გარდა ამისა, ცალკე სალექსიკონო ერთეულად გამოტანილია სიტყვა-სტატია ზმნიზედა, შესაბამისი საილუსტრაციო მასალითა და შესიტყვებით, როგორცაა, **გენეტიკურად მოდიფიცირებული;** მოხსენება შეეხება გენეტიკურად მოდიფიცირებულ ორგანიზმებს (გმო).

## 29. დისკომფორტი

### ქართული ენის განმარტებითი ლექსიკონი

**დისკომფორტი-ი** (დისკომფორტისა) [ლათ. dis-, ბერძ. dys- უარყოფა და ინგ. comfort] წიგნ. ცუდი პირობები (ცხოვრებისა, მუშაობისა, დასვენებისა...)-გაორებული მდგომარეობა ბადებს არასრულფასოვნების კომპლექსებს, შინაგან დისკომფორტს, მუდმივ ოპოზიციონეობის აურას («ლიტ. საქ.»).

### „ქართული ლექსიკონი“

**დისკომფორტი** დისკომფორტის ცხოვრების, მუშაობის, დასვენების ცუდი პირობები, კომფორტის არქონა. დისკომფორტს ვერ ეგუება.

### ქართული ენის ვებკორპუსი

სიხშირული ინდექსი: 961

**დისკომფორტი-ი** (დისკომფორტისა) 1. უხერხული / არახელსაყრელი სიტუაცია, მდგომარეობა; მათგან არახდროს მიგრძენია დისკომფორტი; ერთადერთი დისკომფორტი, რაც მოწვევის შედეგად მიგრძენია, გადაღლილობა; ამის გამო პარკის ვიზიტორებს ექმნებათ სერიოზული დისკომფორტი; მთავრობას უნდა შეეუქმნათ დისკომფორტი, რომ აღარ მოხდეს ადამიანების გათავისუფლება საჯარო სამსახურებიდან; 2. ამა თუ იმ დაავადებისას თანმხლები მდგომარეობა; ჭაზერის გამოყენებით მცირდება სისხლის დანაკარგი, მცირდება პოსტოპერაციული დისკომფორტი; დისონანსის რეზულტატს წარმოადგენს ფსიქოლოგიური დისკომფორტი; ხშირად ამას თან ახლავს დისკომფორტი და მუცლის შებერილობის შეგრძნება.

ქართული ენის განმარტებით ლექსიკონსა და „ქართულ ლექსიკონში“: მოცემული სიტყვის მხოლოდ ერთი მნიშვნელობაა გამოტანილია. რაც შეეხება ქართული ენის ვებკორპუსის მიხედვით შედგენილ სალექსიკონო სიტყვა-სტატიას, მასში გამოიყოფა ორი მნიშვნელობა: 1) უხერხული / არახელსაყრელი სიტუაცია, მდგომარეობა; მათგან არახდროს მიგრძენია დისკომფორტი და 2) ამა თუ იმ დაავადებისას თანმხლები მდგომარეობა; ჭაზერის გამოყენებით მცირდება სისხლის დანაკარგი, მცირდება პოსტოპერაციული დისკომფორტი; დისონანსის რეზულტატს წარმოადგენს ფსიქოლოგიური დისკომფორტი; ხშირად

ამას თან ახლავს დისკომფორტი და მუცლის შებერილობის შეგრძნება. უნდა აღინიშნოს, რომ ზემოთ მოცემული მნიშვნელობები არ არის ასახული ქართული ენის განმარტებით ლექსიკონში.

### 30. ვირუსული

#### ქართული ენის განმარტებითი ლექსიკონი

**ვირუსულ-ი (ვირუსულისა) ვირუს[ებ]ისა, ვირუს[ებ]ის დამახასიათებელი.**

ერთი და იმავე სახის ვირუსული ფორმები და მიკრობების ფორმები, ჩვენი დაკვირვებით, როცა ისინი ერთი-მეორედ გარდაიქცევიან, შეიძლება მიკრობთა და ვირუსთა სხვა სახეებადაც განვითარდნენ («სახალხო განათლა»). ვირუსული დაავადება ვირუსებით გამოწვეული დაავადება. ჩვენ შევძელით მიგველო ვირუსული დაავადების ცხრა აღმძვრელის მიკრობული ფორმები («სახალხო განათლა»).

#### „ქართული ლექსიკონი“

**ვირუსული** ვირუსულის ვირუსით გამოწვეული. ვირუსული გრიპი. მცენარეთა ვირუსული დაავადებანი.

#### ქართული ენის ვებკორპუსი

სიხშირული ინდექსი: 1530

**ვირუსულ-ი (ვირუსულისა) 1. ვირუს[ებ]ისა, ვირუს[ებ]ის დამახასიათებელი;** ცხელების მიზეზად შეიძლება იქცეს ნებისმიერი ვირუსული, ბაქტერიული და პარაზიტული ინფექცია; ვირუსული ანგინა უმეტესად ვითარდება ზედა სასუნთქი გზების რესპირაციული ვირუსული ინფექციის ფონზე; A ჰეპატიტი ღვიძლის ვირუსული ანთების ყველაზე გავრცელებული ფორმაა; **2. პოპულარული/ფართო მასებში ცნობილი;** სამ დღეში უკვე მილიონამდე მნახველი შეაგროვა ვირუსულმა ვიდეომ, რაც ცხადია კარგი მაჩვენებელია; საბალეტო დასის შესრულებით ვირუსულმა ცეკვამ Harlem Shake-მა კლასიკამდეც მიაღწია; ფოტო, რომელიც შეყვარებულისთვის გადაიღო, ინტერნეტში ვირუსულად გავრცელდა; სოციალურ ქსელ ფეისბუქის მეშვეობით შესაძლებელია განწყობის ვირუსულად გავრცელება; ვიდეო ვირუსულად იქცა და 1 კვირაში 9 მილიონი ნახვა დააგროვა; **ვირუსული გრიპი;** ბელა ხუთი თვის ფესხმძიმე იყო, ვირუსული გრიპი რომ შეეყარა; ბოლო მონაცემებით ვირუსული გრიპის ლაბორატო-რიულად

დადასტურებული 439 შემთხვევაა გამოვლენილი; ამ სეზონზე ვირუსული გრიპის A, B ტიპი და H1N1 ვირუსი ლაბორატორიულად 1103 პაციენტს დაუდგინდა; **ვირუსული დაავადება**; ჰეპატიტი არის ვირუსული დაავადება, რომელიც განსაკუთრებულ ყურადღებას მოითხოვს; მეჭვჭი არის კანის ვირუსული დაავადება, შეიძლება კანის ერთ უბანზე მორჩეს, მაგრამ სხვა უბანზე იჩინოს თავი; წითელა ფართოდ გავრცელებული ვირუსული დაავადებაა, რომელიც ადვილად გადადის ჰაერწვეთოვანი გზით; **ვირუსული ინფექცია**; მწვავე რესპირატორული ვირუსული ინფექცია იშვიათად იწვევს ორგანიზმის სერიოზულ დაზიანებას; დაავადების გამწვებ მექანიზმს თითქმის ყოველთვის ვირუსული ინფექცია წარმოადგენს; დელფინების დალუპის მიზეზი ვირუსული ინფექციაა, რომელიც შავ ზღვაში ძუძუმწოვრებში ვრცელდება; **ვირუსული ჰეპატიტი**; ვირუსული ჰეპატიტები მსოფლიოში საკმაოდ გავრცელებული და საშიში დაავადებებია; ვირუსული ჰეპატიტები ორ დიდ ჯგუფად იყოფა; ვირუსული ჰეპატიტების პრობლემა უწინდებურად აქტუალურია.

ქართული ენის განმარტებით ლექსიკონსა და „ქართულ ლექსიკონში“ სალექსიკონო ერთეული **ვირუსული** მხოლოდ ერთი მნიშვნელობითაა წარმოდგენილი. ქართული ენის ვებკორპუსის მიხედვით შედგენილ სიტყვა-სტატიაში კი ასახულია ისეთი უახლესი მნიშვნელობა, როგორცაა: **პოპულარული/ფართო მასებში ცნობილი**; *სამ დღეში უკვე მილიონამდე მნახველი შეაგროვა ვირუსულმა ვიდეომ, რაც ცხადია კარგი მაჩვენებელია; საბალეტო დასის შესრულებით ვირუსულმა ცეკვამ Harlem Shake-მა კლასიკამდეც მიადწია; ფოტო, რომელიც შეყვარებულისთვის გადაიღო, ინტერნეტში ვირუსულად გავრცელდა; სოციალურ ქსელ ფეისბუქის მეშვეობით შესაძლებელია განწყობის ვირუსულად გავრცელება; ვიდეო ვირუსულად იქცა და 1 კვირაში 9 მილიონი ნახვა დააგროვა. გარდა ამისა, მოცემულია ამ სიტყვასთან დაკავშირებული ისეთი შესიტყვებები, როგორებიცაა: **ვირუსული გრიპი**; ბოლო მონაცემებით ვირუსული გრიპის ლაბორატორიულად დადასტურებული 439 შემთხვევაა გამოვლენილი; **ვირუსული დაავადება**; ჰეპატიტი არის ვირუსული დაავადება, რომელიც განსაკუთრებულ ყურადღებას მოითხოვს; **ვირუსული ინფექცია**; მწვავე რესპირატორული ვირუსული ინფექცია იშვიათად იწვევს ორგანიზმის სერიოზულ დაზიანებას; **ვირუსული ჰეპატიტი**; ვირუსული ჰეპატიტები ორ დიდ ჯგუფად იყოფა.*



### 31. ბაშუშეხა

#### ქართული ენის განმარტებითი ლექსიკონი

**ბაშუშეხა (ბაშუშეხისა)** სახელი გააშუქებს, გაშუქდება ზმნათა მოქმედებისა. – 1. **შუქის, ნათელის მოფენა, – განათება.** *ოთახის გაშუქება. – ფიგურ. [ნიკო] ფიქრობდა, რომ... სინათლის მომფენი სოფლის ბნელ ცხოვრებაში სკოლა იყო და ამ სინათლის მოფენა, ამ სიბნელის გაშუქება საბედნიერო საქმედ მიაჩნდა (ეკ. გაბ.).* 2. **გადატ. გარჩევა, განხილვა, განმარტება; ნათელყოფა, წარმოჩენა.** *ახალი პრობლემის გაშუქების დროს გამოირკვა, რომ... აქამდე არსებული თეორიებისა და გაბატონებულ შეხედულებათა გათვალისწინება და კრიტიკული განხილვაა საჭირო (ივ. ჯავახ.). ვახტანგმა შენიშნა, რომ «ვეფხისტყაოსნის» ზოგიერთი ადგილი... გარკვევას საჭიროებდა და გაშუქებას მოითხოვდა (ლ. მენ.).* 3. **სპეც. რენტგენის სხივებით გასინჯვა, შემოწმება (ცოცხალი ორგანიზმისა).** *არადა, დროდადრო ადამიანს აუცილებლად სჭირდება რენტგენით გაშუქება («დრონი»).*

#### „ქართული ლექსიკონი“

**ბაშუშეხა** გაშუქების (გააშუქებს, გაიშუქებს, გაშუქდება, გააშუქებიან) 1. **მზის, ლამპისა და მისთ. შუქის სხივების მოფენა.** *გზა ფარნით გააშუქა.* 2. **გადატ. განხილვა. ახსნა.** *მომხსენებელმა საკითხი კარგად გააშუქა. პრეზიდენტის ვიზიტი პრესაში გაშუქდა.* 3. **სამედიცინო გამოკვლევა რენტგენის სხივებით.** *ფილტვები გაუშუქა. გაშუქებული. რაც გააშუქეს. ტელეეზით გაშუქებული ფაქტი.*

#### ქართული ენის ვებკორპუსი

სისწირული ინდექსი: 1089

**ბაშუშეხა (ბაშუშეხისა)** 1. **მედიასაშუალებებით ამა თუ იმ ინფორმაციის გადაცემა;** *მოჭიდავეების აქციის პირდაპირ ეთერში გაშუქება; ადგილობრივი მთავრობის აქტივობების საინფორმაციო გამოშვებებისათვის უჩვეულოდ დიდი დროით გაშუქება; უამრავმა მშობელმა გააპროტესტა ამ კლიპის ტელევიზიით გაშუქება;* 2. **გარჩევა, განხილვა, განმარტება;** *შეუფერებელ დროს მოგვიწია ამ ფაქტის საჯაროდ გამოტანა და გაშუქება; ერთი წერილით შეუძლებელია, და არასასურველიც, პრობლემათა მთელი სპექტრის გაშუქება;* 3. **სპეც. რენტგენის**

სსიფებით გასინჯვა, შემოწმება (ცოცხალი ორგანიზმისა); ყოველ მისვლაზე მიწვევს სპიროგრაფიის შემოწმება და ფილტვების გაშუქება; რენტგენით გაშუქებას და რენტგენოგრაფიას დიაგნოზის დასასმელად იყენებენ; 4. განათება; ყველა შენობის კედლების სპეციალური რენტგენული აპარატურით გაშუქება, მოგესხენებათ, შეუძლებელია.

ქართული ენის განმარტებით ლექსიკონსა და „ქართულ ლექსიკონში“ აღნიშნული სიტყვის სამი მნიშვნელობა გამოიყოფა. რაც შეეხება ქართული ენის ვებკორპუსის მიხედვით შედგენილ სალექსიკონო სიტყვა-სტატიას, აქ ოთხი მნიშვნელობა გამოვყავით, მათ შორის ერთი მნიშვნელობა, რომელიც არ არის ასახული ქართული ენის განმარტებით ლექსიკონში: **მედიასაშუალებებით ამა თუ იმ ინფორმაციის გადაცემა; მოჭიდავეების აქციის პირდაპირ ეთერში გაშუქება; დგილობრივი მთავრობის აქტივობების საინფორმაციო გამოშვებებისათვის უჩვეულოდ დიდი დროით გაშუქება; უამრავმა მშობელმა გააპროტესტა ამ კლიპის ტელევიზიით გაშუქება.**

## 32. პროგრესული

### ქართული ენის განმარტებითი ლექსიკონი

**პროგრესულ-ი (პროგრესულია)** წიგნ. პროგრესის ხელის შემწყობი, – პოლიტიკურად და საზოგადოებრივად მოწინავე. პროგრესული კლასი. – მთელი პროგრესული კაცობრიობა ადევნებს თვალყურს საბჭოთა ხალხის შრომითს სავშირო საქმეებს ომისშემდგომ წლებში («კომ.»). პროგრესული ასიმილაცია ლინგვ. იხ. ასიმილაცია.

### „ქართული ლექსიკონი“

**პროგრესული** პროგრესულის მოწინავე, განვითარებული, პროგრესის ხელშემწყობი. პროგრესული ახალგაზრდა. პროგრესული იდეა. შეცდომაა. პროგრესიული.

### ქართული ენის ვებკორპუსი

სისშირული ინდექსი: 1058

**პროგრესულ-ი (პროგრესულია)** პროგრესის შემწყობი და იდეის გამტარებელი; ეს გუნდი ნამდვილად იყო უფრო პროგრესული, ვიდრე

ნებისმიერი სხვა; პროგრესული , დასავლური ღირებულებებზე ორიენტირებული ხელისუფლება; მსოფლიოში მყარი წესრიგი მხოლოდ იმ შემთხვევაში იქნება, თუ უფრო პროგრესული , ევროპული ხედვა გაიმარჯვებს; **პროგრესული აზროვნება;** ცდილობენ, ყველას დაუმტკიცონ თავიანთი უკომპლექსობა და პროგრესული აზროვნება; მისი პროგრესული აზროვნება თვალში ეკლად ესობოდა მეფის რუსეთის ხელისუფლებას; მიუხედავად პერსონაჟების პროგრესული აზროვნებისა, ისინი მაინც ჩაკეტილ სივრცეში ცხოვრობენ; **პროგრესული განვითარება;** მან განსაზღვრა ევროპული ფილოსოფიის შემდგომი პროგრესული განვითარება; რევოლუცია ძალზე იშვიათად ყოფილა საზოგადოების პროგრესული განვითარების წყარო; ამ ყველაფერს კი ემატება მსოფლიო ცივილიზაციის პროგრესული განვითარება, სადაც ადამიანი უფუნქციოდ რჩება; დროა ძალიან რადიკალური და მტკივნეული ცვლილებების, რომელმაც უნდა უზრუნველყოს ქვეყნის და საზოგადოების პროგრესული განვითარება; **პროგრესული იდეა;** მე მხოლოდ და მხოლოდ მოწინავე და პროგრესული იდეებით მოვდივარ სამინისტროში; ეს მართლაც დემოკრატიული და პროგრესული იდეების გამარჯვებაა; სხვადასხვა სფეროებში ნოვატორული პროექტებისა და პროგრესული იდეების შემუშავება; **პროგრესული ტექნოლოგია;** წვრილ მეურნეობებში ვერ ხერხდება პროგრესული ტექნოლოგიების დანერგვა; მემბრანული ტექნოლოგია შედის მსოფლიოს პროგრესული ტექნოლოგიების ხუთეულში; საფრთხო სავაჭრო სერვისი ასევე შეიძლება გახდეს მნიშვნელოვანი ფაქტორი ინვესტიციების და პროგრესული ტექნოლოგიების მისაზიდად.

როგორც ქართული ენის განმარტებით ლექსიკონში, „ქართულ ლექსიკონში“, ასევე ქართული ენის ვებკორპუსის მიხედვით შედგენილ სიტყვა-სტატიაში მოცემული სიტყვის ერთი მნიშვნელობა გამოიყოფა, განსხვავება მდგომარეობს, საინსტრუქციო მასალის მრავალფეროვნებასა და იმ შესიტყვებებში, რაც ქართული ენის ვებკორპუსის მიხედვით შედგენილ სტატიაშია ასახული. ეს შესიტყვებებია: **პროგრესული აზროვნება;** ცდილობენ, ყველას დაუმტკიცონ თავიანთი უკომპლექსობა და პროგრესული აზროვნება; **პროგრესული განვითარება;** მან განსაზღვრა ევროპული ფილოსოფიის შემდგომი პროგრესული განვითარება; **პროგრესული იდეა;** მე მხოლოდ და მხოლოდ მოწინავე და პროგრესული იდეებით მოვდივარ სამინისტროში; **პროგრესული**

ტექნოლოგია; წვრილ მეურნეობებში ვერ ხერხდება პროგრესული ტექნოლოგიების დანერგვა.

### 33. მარბალიტი

#### ქართული ენის განმარტებითი ლექსიკონი

**მარბალიტი-ი (მარბალიტისა)** 1. მარცვლისებური ძვირფასი თვალი თეთრი, შავი ან ყვითელი ფერისა; წარმოიშვება ზოგიერთი მოლუსკის სხეულში; ხმარობენ სამკაულად. მეც მიბოძეს მათ საკუთრად საჩუქარი დიდძალით: სამკაული, ოქრო, ვერცხლი, მარგალიტი და თვალი (აკაკი). // ფიგურ. [მდინარე] სცემდა ნაპირებს, ისროდა აქეთ-იქით მარგალიტ წინწკლებს (ნ. ლომ). 2. გადატ. მეტისმეტად კარგი, არაჩვეულებრივი, ძვირფასი, საუკეთესო რამ. ბევრი არა უწერია რა [ი. დავითაშვილს], მაგრამ რაც დაწერა, იმათში ასეთი მარგალიტები ამოირჩევა, რომ კაცს დაატკობს აზრითაც და სიტყვიერებითაც (ილია). «ნუ გეშინიანთ, ჩემო შვილებო, ტკბილებო, მარგალიტებო! (ვაჟა). **ობოლი მარგალიტი** დიდი მარგალიტი, რომელიც იშვიათია და ძვირფასად ითვლება. // გადატ. უბადლო, იშვიათი, რჩეული. ბევრმა კარგად აშენებულმა გლეხმა სთხოვა [პეპიას] ქალი, მაგრამ არავის მიათხოვა, თითქო სხვისათვის არ ემეტებოდა ის ობოლი მარგალიტი (ილია). მაგისტანა [შექსპირივით, ბაირონსავით] ობოლ მარგალიტებს ძალიან იშვიათად მოჰყრის ხოლმე მდინარე ცხოვრებისა (ილია). 2. რთული ფუძის მეორე შემადგენელი ნაწილი. თვალ-მარგალიტი.

„ქართული ლექსიკონი“

**მარბალიტი** მარგალიტის ძვირფასი ქვა, მინერალური წარმონაქმნი, რომელიც ვითარდება ზოგიერთი მოლუსკის ნიჟარაში მოხვედრილ უცხო სხეულზე სადაფის შემოზრდით. ზღვის მარგალიტი. შავი მარგალიტი. მარგალიტის მძივი.

ქართული ენის ვებკორპუსი:

სიხშირული ინდექსი: 559

**მარბალიტი-ი (მარბალიტისა)** 1. მარცვლისებური ძვირფასი თვალი თეთრი, შავი ან ყვითელი ფერისა; წარმოიშვება ზოგიერთი მოლუსკის სხეულში;

**ხმარობენ სამკაულად;** ერთადერთი ძვირფასი თვლებიდან არ მიყვარს მარგალიტი , ზედმეტად ქაღურია; ნახევრადძვირფასი თვლებისგან და მასალებისგან არის გაკეთებული, როგორცაა: მარგალიტი, ფირუზი; ძალიან უყვარს მარგალიტის ყელსაბაში, განსაკუთრებით კი შავი მარგალიტი; **2. მეტისმეტად კარგი, არაჩვეულებრივი, ძვირფასი, საუკეთესო რამ;** თავისს ამხანავ გოგოებში მარო, როგორც მარგალიტი, ისე ირჩეოდა; შენი მეგობრები შენი სიმდიდრეა, შენი მარგალიტი, ოქრო და ბრილიანტი; ქილილა და დამანა მსოფლიო ლიტერატურის ძვირფასი მარგალიტო; **3. ჩვეულ. მრავლობ.** **სასაცილო / სამარცხვინო ამბების, ინფორმაციის და მისთ. კრებული;** ჯერ სადა ვართ, წინ კიდე ახალ-ახალი მარგალიტები გველის ალბათ; სოციალურ ქსელებს საკმაოდ ჰყავს მარგალიტები და ამ მარგალიტების რაოდენობა ისე იზრდება; შეკრიბე შენი მარგალიტები და წიგნად გამოეცო; **ობოლი მარგალიტი დიდი მარგალიტი, რომელიც იშვიათია და ძვირფასად ითვლება;** გადატ. **შეუდარებელი, ძვირფასი რამ;** თითქო სხვისათვის არ ემეტებოდა ის ობოლი მარგალიტი; ამ სასმისით მეგობრებო გულთა შორის გავდე ხიდი და სამშობლო დაგილოცეთ, ეს ობოლი მარგალიტი; მთელ ქორწილის ხალხში ისე ვირჩეოდი, როგორც ობოლი მარგალიტი.

ქართული ენის განმარტებით ლექსიკონში აღნიშნული სიტყვის ორი მნიშვნელობაა გამოტანილი. „ქართულ ლექსიკონში“ კი ამავე სიტყვის ერთი მნიშვნელობაა მოცემული. ქართული ენის ვებკორპუსის მიხედვით კი მესამე მნიშვნელობაც გამოვყავით. ეს მნიშვნელობაა: **სასაცილო / სამარცხვინო ამბების, ინფორმაციის და მისთ. კრებული;** ჯერ სადა ვართ, წინ კიდე ახალ-ახალი მარგალიტები გველის ალბათ; სოციალურ ქსელებს საკმაოდ ჰყავს მარგალიტები და ამ მარგალიტების რაოდენობა ისე იზრდება; შეკრიბე შენი მარგალიტები და წიგნად გამოეცი.

### 34. ჰუმანიტარული

#### ქართული ენის განმარტებითი ლექსიკონი

**ჰუმანიტარულ-ი** (ჰუმანიტარულისა) ლათ. [humanitas ადამიანურობა] ადამიანის (და კულტურის) შესახები (მეცნიერება...). ჰუმანიტარული მეცნიერება. – ჰუმანიტარული განათლება. – მეთერთმეტეკლასელები მარტო ჰუმანიტარულ საგნებში როდი იჩენენ წარმატებებს («კომ.»).

#### „ქართული ლექსიკონი“

**ჰუმანიტარული** ჰუმანიტარულის 1. ხალხის კულტურისა და ისტორიის შესწავლასთან დაკავშირებული. ჰუმანიტარული ფაკულტეტი. ჰუმანიტარული განათლება. 2. რასაც ჰუმანურობა უდევს საფუძვლად, ადამიანის კეთილდღეობისკენ მიმართული. ჰუმანიტარული ორგანიზაცია. ჰუმანიტარული დახმარება. შეცდომაა გუმანიტარული.

#### ქართული ენის ვებკორპუსი

სისწირული ინდექსი: 1059

**ჰუმანიტარულ-ი** (ჰუმანიტარულისა) 1. ადამიანისა და კულტურის შესახებ მეცნიერებებთან დაკავშირებული; ოჯახში ჰუმანიტარული განათლების არავინ მყოლია; ჰუმანიტარული განხრით უფრო ძლიერი ვიყავი; 2. ადამიანური და კულტურული; ყველაფერს აკეთებთ, რათა ხელი შეუწყოთ ჰუმანიტარული, კულტურული და სულიერი კონტაქტების შენარჩუნებას მოძმე ერებს შორის; ჩვენი საქმიანობის ძირითადი პრინციპები ჰუმანიტარული აქტივობებით შემოიფარგლება; ინტერვენცია განსხვავებულად, ჰუმანიტარული გადმოსახედიდან შეფასდება; 3. ადამიანის დახმარებასთან დაკავშირებული; გამოკვლევები მოსახლეობას ჰუმანიტარული აქციის ფარგლებში ჩაუტარდა; 31-ე ბატალიონმა მოსახლეობას ჰუმანიტარული და სამედიცინო დახმარებაც აღმოუჩინა; დარწმუნებულია, რომ ჰუმანიტარული და ფინანსური დახმარება ღარიბულ სახელმწიფოებში ფუჭად არ იხარჯება; ჰუმანიტარული დახმარება; კომპაქტურად ჩასახლებულმა დევნილებმა ჰუმანიტარული დახმარებაც მიიღეს; მათ არ გააჩნიათ არც ჰუმანიტარული დახმარება, არც სამუშაო ადგილები; დანიშნულებისამებრ, მიეწოდება მედიკამენტები, გაეწევა ჰუმანიტარული დახმარება; ჰუმანიტარული კატასტროფა; პოლიტიკური კრიზისის გარდა იემენი

ჰუმანიტარული კატასტროფის ზღვარზეც დგას; ექსტრემისტების გაძლიერება ჰუმანიტარული კატასტროფის საფრთხის შემცველია; უდიდესი გაჭირვების მიუხედავად, მასობრივი შიმშილის და ამის შედეგად გამოწვეული ჰუმანიტარული კატასტროფის მოწმენი არ გავმხდარვართ; **ჰუმანიტარული კრიზისი**; გაეროს შეფასებით, სირიის კონფლიქტი ყველაზე დიდი ჰუმანიტარული კრიზისია; დღესდღეობით არ არსებობს ჰუმანიტარული კრიზისის რაიმე მტკიცებულება; ტრიპოლში დღეს გაეროს დელეგაცია ჩავა და ჰუმანიტარული კრიზისის მასშტაბებს შეაჯამებს; **ჰუმანიტარული მეცნიერებები**; მამაკაცები ძლიერები არიან ზუსტ მეცნიერებებში, ქალებს კი უკეთ გამოხდით ჰუმანიტარული მეცნიერებები; შემდეგ დღეებში სტუდენტები მუშაობდნენ ორ ჯგუფად - ჰუმანიტარული მეცნიერებისა და სოციალური მეცნიერების მიმართულებით; დღეს საქართველოში ჰუმანიტარული მეცნიერების წარმომადგენლებს თამადა შეუძლიათ დამოუკიდებელი მეცნიერული კვლევა აწარმოონ; **ჰუმანიტარული მისია**; არის საგანგებო რეაგირების მიმართულება, ჰუმანიტარული მისია, რომელიც კრიზისულ ვითარებაში მუშაობს; ჩვენ ასევე ხაზგასმით ავლნიშნავთ ამ ტერიტორიებზე ჰუმანიტარული მისიების დაშვების აუცილებლობას; წლების განმავლობაში იღებდნენ ჰუმანიტარული მისიის ტვირთს, რომლის რეალიზებას ახდენდნენ ადგილობრივ ბაზარზე; **ჰუმანიტარული ორგანიზაცია**; საერთაშორისო ჰუმანიტარული ორგანიზაციები აქტიურად მუშაობენ, რათა თავიდან აიცილონ ეპიდემია; ჰუმანიტარული ორგანიზაციების თქმით, შეიდი მილიონი ადამიანი შიმშილის ზღვარზე იმყოფება; რეგიონში ჰუმანიტარულ ორგანიზაციებს შეუფერხებელი მუშაობის შესაძლებლობა უნდა ჰქონდეთ; **ჰუმანიტარული პროგრამა**; ანტირაბიული აცრების ჰუმანიტარული პროგრამა; დაკავებულნი არიან ცოცხალი, პრაქტიკული ქრისტიანობით, სოციალური და ჰუმანიტარული პროგრამებით; თხოვნით მიმართა ფრანგულ მხარეს ამ მიმართულებით ჰუმანიტარული პროგრამების განხორციელებაზე; **ჰუმანიტარული პროექტი**; განაგრძობენ მჭიდრო თანამშრომლობასა და კოორდინაციას ჰუმანიტარული პროექტების განხორციელების პროცესში; მნიშვნელოვანია ჰუმანიტარული პროექტების განხორციელების ხელშეწყობა კოორდინირების გზით; ლიტვის ელჩმა ჰუმანიტარული პროექტების განხორციელებისათვის მზადყოფნა გამოთქვა; **ჰუმანიტარული საგნები**; მე უფრო ჰუმანიტარული საგნები მიზიდავდა; მე პატარა ასაკიდანვე ჰუმანიტარული საგნებისკენ ვიყავი მიდრეკილი; ამაში

ჰუმანიტარული საგნების როლი ძალიან დიდია; **ჰუმანიტარული სტატუსი**; ლტოლვილის სტატუსის გარდა, კანონმდებლობაში ჰუმანიტარული სტატუსის მატარებელი პირის ცნებაც შემოდის; ორგანიზაციამ საიმპორტო საქონლისათვის ჰუმანიტარული სტატუსის მინიჭების მიზნით უნდა შეავსოს განაცხადი; კანონპროექტში ლტოლვილის და ჰუმანიტარული სტატუსის მქონე პირთა ოჯახის წევრების დაცვა უკეთესად არის გათვალისწინებული; **ჰუმანიტარული ტვირთი**; მადლობას ვუხდით იმ ორგანიზაციას, რომელმაც ჰუმანიტარული ტვირთი გამოგვიგზავნა; ქალთა ორგანიზაციამ ომიდან დაზარალებულებისათვის შეაგროვა ჰუმანიტარული ტვირთი; დაზარალებულ მოსახლეობას ჰუმანიტარული ტვირთი მიაწოდეს; **ჰუმანიტარული ფაკულტეტი**; ბაკალავრიატის დასრულების შემდეგ მან აქვე ჰუმანიტარული ფაკულტეტის მაგისტრატურაში ჩააბარა; დავამთავრე ჰუმანიტარული ფაკულტეტი ფსიქოლოგიის განხრით; ივანე ჯავახიშვილის უნივერსიტეტის ჰუმანიტარული ფაკულტეტი.

ქართული ენის განმარტებით ლექსიკონში მოცემულ სალექსიკონო ერთეულისათვის გამოტანილია მხოლოდ ერთი მნიშვნელობა. ხოლო „ქართულ ლექსიკონში“ მოცემული ამავე სიტყვის ორი მნიშვნელობა. ქართული ენის ვებკორპუსის მიხედვით შედგენილ სიტყვა-სტატაში კი სამი მნიშვნელობა გამოვყავით. ეს მნიშვნელობებია: **1. ადამიანისა და კულტურის შესახებ მეცნიერებებთან დაკავშირებული**; ოჯახში ჰუმანიტარული განათლების არავინ მყოლია; **2. ადამიანური და კულტურული**; ველაფერს აკეთებთ, რათა ხელი შეუწყოთ ჰუმანიტარული, კულტურული და სულიერი კონტაქტების შენარჩუნებას მოძმე ერებს შორის და **3. ადამიანის დახმარებასთან დაკავშირებული**; გამოკვლევები მოსახლეობას ჰუმანიტარული აქციის ფარგლებში ჩაუტარდა. გარდა ამისა, გამოვყავით, ისეთი შესიტყვებები, როგორებიცაა: **ჰუმანიტარული დახმარება**; კომპაქტურად ჩასახლებულმა დევნილებმა ჰუმანიტარული დახმარებაც მიიღეს; **ჰუმანიტარული კატასტროფა**; პოლიტიკური კრიზისის გარდა იემენი ჰუმანიტარული კატასტროფის ზღვარზეც დგას; **ჰუმანიტარული კრიზისი**; გაეროს შეფასებით, სირიის კონფლიქტი ყველაზე დიდი ჰუმანიტარული კრიზისია; **ჰუმანიტარული მეცნიერებები**; მამაკაცები ძლიერები არიან ზუსტ მეცნიერებებში, ქალებს კი უკეთ გამოსდით ჰუმანიტარული მეცნიერებები; **ჰუმანიტარული მისია**; არის საგანგებო



რეაგირების მიმართულება, ჰუმანიტარული მისია, რომელიც კრიზისულ ვითარებაში მუშაობს; ჰუმანიტარული ორგანიზაცია; საერთაშორისო ჰუმანიტარული ორგანიზაციები აქტიურად მუშაობენ, რათა თავიდან აიცილონ ეპიდემია; ჰუმანიტარული პროგრამა; ანტირაბიული აცრების ჰუმანიტარული პროგრამა; ჰუმანიტარული პროექტი; განავრცობენ მჭიდრო თანამშრომლობასა და კოორდინაციას ჰუმანიტარული პროექტების განხორციელების პროცესში; ჰუმანიტარული საგნები; მე უფრო ჰუმანიტარული საგნები მიზიდავდა; ჰუმანიტარული სტატუსი; ლტოლვილის სტატუსის გარდა, კანონმდებლობაში ჰუმანიტარული სტატუსის მატარებელი პირის ცნებაც შემოვიდა; ჰუმანიტარული ტვირთი; მადლობას ვუხდით იმ ორგანიზაციას, რომელმაც ჰუმანიტარული ტვირთი გამოგვიგზავნა; ჰუმანიტარული ფაკულტეტი; ბაკალავრიატის დასრულების შემდეგ მან აქვე ჰუმანიტარული ფაკულტეტის მაგისტრატურაში ჩააბარა.

### 35. ვირუსი

#### ქართული ენის განმარტებითი ლექსიკონი

**ვირუსი-ი (ვირუსისა)** [ლათ. virus შხამი] მედიც. ინფექციურ დაავადებათა გამომწვევი – მიკროსკოპით უხილავი – ცილოვანი არსება. თანამედროვე მეცნიერება ვირუსებს –უმცირეს, მიკროსკოპით უხილავ ცილოვან არსებებს, მკვეთრად განასხვავებს მიკრობებისაგან – უჯრედოვანი ორგანიზმებისაგან («სახალხო განათლ.»).

#### „ქართული ლექსიკონი“

**ვირუსი ვირუსის 1.** გადამდები დაავადების გამომწვევი მიკროორგანიზმი. ვირუსის საწინააღმდეგო ვაქცინა. **2.** კომპიუტერული პროგრამა, რომელიც მრავლდება და აზიანებს კომპიუტერის პროგრამულ ან აპარატულ ნაწილს. ვირუსმა ინფორმაცია წაშალა. ვირუსული. ვირუსით გამოწვეული. ვირუსული გრიპი. მცენარეთა ვირუსული დაავადებანი.

#### ქართული ენის ვებკორპუსი

სიხშირული ინდექსი: 1530

**30რუს-0 (30რუსისა) 1. მედიც. ინფექციურ დაავადებათა გამომწვევი – მიკროსკოპით უხილავი – ცილოვანი არსება; ანტიბიოტიკები ვირუსებზე ზემოქმედებას ვერ ახდენენ; ციტრუსებში შემავალი ვიტამინები ორგანიზმს აძლიერებს და ვირუსებთან ბრძოლაში ეხმარება; სიცხე ორგანიზმის ვირუსების წინააღმდეგ ბრძოლის გამოსატყულებათ; 2. კომპ. სპეციალური პროგრამა, რომელიც აზიანებს კომპიუტერულ სისტემას; კომპიუტერული პროგრამების ფართო განვითარებას მოჰყვა ჰაკერებისა და ახალი კომპიუტერული ვირუსების გამავრცელებელთა მთელი პლეადის გაჩენა; კომპიუტერებში შეაღწია ვირუსმა , რომელსაც ვიდეოთვალივალის და ხმის ჩაწერის ფუნქცია ჰქონდა; ამ ვირუსმა გასულ წელს 4 მილიონზე მეტი კომპიუტერი დააზიანა მსოფლიოს მასშტაბით.**

ქართული ენის ლექსიკონში სალექსიკონო ერთეული **ვირუსი** მხოლოდ ერთი მნიშვნელობითაა გამოტანილი, ხოლო „ქართულ ლექსიკონში“ ორი მნიშვნელობაა მოცემული. რაც შეეხება ქართული ენის ვებკორპუსის მიხედვით შედგენილ სიტყვა სტატიას, მასში ორი მნიშვნელობა გამოვყავით, მათ შორის ისეთი, რომელიც არ არის ასახული ქართული ენის განმარტებით ლექსიკონში: **კომპ. სპეციალური პროგრამა, რომელიც აზიანებს კომპიუტერულ სისტემას; კომპიუტერული პროგრამების ფართო განვითარებას მოჰყვა ჰაკერებისა და ახალი კომპიუტერული ვირუსების გამავრცელებელთა მთელი პლეადის გაჩენა; კომპიუტერებში შეაღწია ვირუსმა, რომელსაც ვიდეოთვალივალის და ხმის ჩაწერის ფუნქცია ჰქონდა; ამ ვირუსმა გასულ წელს 4 მილიონზე მეტი კომპიუტერი დააზიანა მსოფლიოს მასშტაბით.**

\*\*\*

ამგვარად, მეოთხე თავში განვიხილეთ ქართული ენის ვებკორპუსი - KaWaC, როგორც წყარო ქართულ-ევროპული ტიპის სასწავლო ლექსიკონების სიტყვა-სტატიების შესადგენად. ამ მიზნით წარმოვადგინეთ სადისერტაციო ნაშრომის კიდევ ერთი სიახლე – ქართული ენის სიხშირული „ენ-გრამები“, რაც აღნიშნავს მომიჯნავე / მოსაზღვრე / მეზობელი ფონემების, სიტყვების, სიტყვათა წყვილების ერთად რეალიზების სიხშირეს. აქ „ენ-გრამებს“ განვიხილავთ მომიჯნავე სიტყვების ერთად რეალიზების მიხედვით, სადისერტაციო ნაშრომში საილუსტრაციოდ მოვიყვანეთ რამდენიმე მაგალითი ორკომპონენტიანი, სამკომპონენტიანი, ოთხკომპონენტიანი და ხუთკომპონენტიანი

„ენ-გრამებისა“. რაც შეეხება უშუალოდ „ენ-გრამების“ სრულ ცხრილს, მისი მოცულობიდან გამომდინარე, იგი ელექტრონულად აქვს დართული წინამდებარე სადოქტორო ნაშრომს.

წინამდებარე თავში საილუსტრაციოდ წარმოვადგინეთ ჩვენ მიერ შედგენილი სიტყვა-სტატიების ნაწილი და აღნიშნული სიტყვა-სტატიები შევადარეთ ქართული ენის განმარტებითს ლექსიკონისა და „ქართული ლექსიკონის“ შესაბამის სიტყვა-სტატიებს. როგორც შედარებამ აჩვენა, ქართული ენის ვებკორპუსის მიხედვით შედგენილი სიტყვა-სტატიები ბევრად ინფორმაციულია და მოიცავს ისეთ ახალ მნიშვნელობებსა და რეალიებს, რაც ქართული ენის ლექსიკონებში არ არის ასახული და ვფიქრობთ, რომ მათი ასახვა სასწავლო ტიპის ლექსიკონებში აუცილებელია.

ვებკორპუსის ბაზაზე შედგენილი სიტყვა-სტატიები ასევე მდიდარია თანამედროვე ქართულ ენაში გავრცელებული შესიტყვებებითა და იდიომებით. ჩატარებულმა ანალიზმა ცხადყო, რომ ჩვენ მიერ შექმნილი ქართული ენის ვებკორპუსი მეტად ეფექტურია ქართული სალექსიკონო სიტყვა-სტატიების შესადგენად, ყველაზე ხშირად ხმარებული სიტყვების პოლისემიური მნიშვნელობების, შესიტყვებებისა და იდიომების გამოსავლენად.

## დასკვნა

სადისერტაციო ნაშრომი, როგორც ეს შესავალშიც ითქვა, არის ქართულ ლექსიკოგრაფიაში კორპუსული მეთოდოლოგიის გამოყენების პირველი მცდელობა. სადოქტორო კვლევის ძირითადი სიახლე და შედეგია ლიდსის უნივერსიტეტში შექმნილი 150-მილიონსიტყვიანი **ქართული ენის ვებკორპუსი** და აღნიშნული ვებკორპუსის მიხედვით დადგენილი **ქართული ენის სინშირული ლექსიკა** – 10,000 ყველაზე ხშირად ხმარებული ქართული სიტყვა. აღნიშნული სიტყვიანი გახდება ქართულ-ევროპული ტიპის ახალი სასწავლო ლექსიკონების სიტყვიანის განსაზღვრის საფუძველი. ჩატარებულმა კვლევამ ასევე გამოავლინა ჩვენ მიერ შექმნილი ქართული ენის ვებკორპუსის ეფექტურობა აღნიშნული ლექსიკონების ქართული ნაწილის სალექსიკონო სიტყვა-სტატიების შესადგენად.

სადისერტაციო ნაშრომში „ქართულ-ევროპული ტიპის ახალი სასწავლო ლექსიკონების შედგენის საკითხები კორპუსული მეთოდოლოგიის გამოყენებით“ დეტალურადაა აღწერილი ჩატარებული კვლევის თითოეული ეტაპი და გამოყენებული მეთოდოლოგია:

1. ქართული ენის ვებკორპუსი მოიცავს 150 მილიონ სიტყვას, უფრო ზუსტად:
  - ტექსტი: 618468 ვებგვერდი - 697 ვებსაიტიდან
  - სიტყვა: 182845341, 2865042 ტიპი
  - ტოკენი: 23550807, 3447266 ტიპი
2. ქართული ენის ვებკორპუსში რეპრეზენტაციულობის მიღწევის მიზნით შეირჩა 697 ვებსაიტი, რაც მოტიანობაში 618468 ვებგვერდს შეადგენს.
3. ვებგვერდები პროგრამა wget-ის (World Wide Web და get) გამოყენებით ჩამოიტვირთა.
4. უშუალოდ ვებგვერდების ჩამოტვირთვამდე განისაზღვრა 300 ფუნქციური სიტყვა (function words) ქართული ენისათვის, რომელიც შეირჩა დევიდ თაგუელის ქართული ენის კორპუსის მიხედვით. ფუნქციური მნიშვნელობის მქონე ქართული სიტყვიანის შერჩევა კორპუსული მასალის შეგროვების პროცესში მნიშვნელოვანია იყო იმისათვის, რომ გამოგვერიცხა ვებგვერდებიდან ტექსტური მასალის ჩამოტვირთვამდე არაქართული

გვერდები, ანიმაციები და სხვა ლინგვისტურად არარელევანტური ინფორმაცია. ქართული სიტყვის წარმოდგენა ამ ნაშრომის ერთ-ერთ სიახლეს წარმოადგენს.

5. ჩამოტვირთული მასალის ლინგვისტური დამუშავება სპეციალური კორპუსული ინსტრუმენტის – BootCatTools-ის საშუალებით განხორციელდა. ვებრესურსების ლინგვისტური დამუშავებისათვის ასევე დავეყრდენით, ე.წ. „კლინივალის“ (CleanEval: cleaning webpages) გამოცდილებას (Baroni, Chantree, Kilgarriff, Sharoff 2008). „კლინივალი“ არის შეფასება ვებგვერდების გასუფთავების კუთხით და მიზნად ისახავს ვებმონაცემების კორპუსულ რესურსად გამოყენებას. უნდა აღინიშნოს, რომ ჩვენ მიერ ვებგვერდების გასუფთავებაში გამოყენებული მეთოდოლოგია არის უახლესი და არის ლიდის უნივერსიტეტის (დიდი ბრიტანეთი), ტრენტის უნივერსიტეტისა (იტალია) და „ლექსიკალ კომპიუტინგის“ (Lexical Computing LTD) ერთობლივი პროექტი.
6. ქართული ენის ვებკორპუსის ვებგვერდების ანოტირება და გასუფთავება მოიცავდა შემდეგ პუნქტებს:
  - 1) HTML-ისა და ჯავასკრიპტების (JavaScript) კოდების მოხსნას, ლინგვისტურად არარელევანტური ინფორმაციის წაშლას (კოლონტიტულები, საავტორო უფლების აღმნიშვნელი ნიშნები, ბმულები, ცალკეულ ვებგვერდებში გამეორებული მასალები და ა.შ.);
  - 2) მინიმალური სიმბოლოების გამოყენებით ძირითადი კოდების მინიჭება (იხილეთ სურათი 3) ტექსტების სათაურებისათვის, პარაგრაფებისათვის და ა.შ.
7. ამჟამად KaWaC განთავსებულია ლიდის უნივერსიტეტის ვებგვერდზე და აღჭურვილია ონლაინ საძიებო სისტემით, ამ ეტაპზე შესაძლებელია სიტყვანის (Word List), კონკორდანსების ძიება კორპუსში და შესაბამისი კონტექსტის ნახვა. კორპუსში შესაძლებელია ძიება მეტყველების ნაწილების მარკერების – POS Tag (Part-of-Speech Tag) მიხედვით. გარდა ამისა, ჩვეულებრივი (არა მეტყველების ნაწილების მარკერების) ძიების

შემთხვევაშიც გამოტანილია ამა თუ იმ სიტყვის შესაბამისი მეტყველების ნაწილის მარკერი და ლემა.

8. ვებკორპუსის დაბალანსება დაეფუძნა „ფუნქციური უანრული განზომილებების“ (Functional Genre Dimensions) მეთოდს. გამოიყო შემდეგი უანრები: პრესა, ახალი ამბები, პირადი ბლოგები, ფორუმები და სხვ.
9. შეიქმნა ქართული ენის მეტყველების მარკერების სისტემა, რომლის პლატფორმად გამოყენებულია „მულტიტექსტური მორფოსინტაქსური პარამეტრები“ (MULTEXT-East Morphosyntactic Specifications - MTE), მეოთხე ვერსია. ნაშრომში სრული სახითაა მოცემული მეტყველების ნაწილების მარკერების, ე.წ. ტაგსეტების სისტემა. ქართული ენისათვის მეტყველების ნაწილების მარკერების სისტემების „მულტიტექსტური მორფოსინტაქსური პარამეტრების“ მეოთხე ვერსიის შესაბამისად შემუშავება მნიშვნელოვანია იმ თვალსაზრისითაც, რომ ქართული ენის ვებკორპუსი შეიძლება შევადაროთ მორფოსინტაქსური პარამეტრების ბაზაში მოცემული 16 ენის კორპუსებს.
10. ნაშრომში ასევე აღწერილია ქართული ენის ვებკორპუსის ავტომატური მორფოლოგიური ანოტირების პროცესი, კერძოდ, საწვრთნელი მასალის მომზადება, მორფოლოგიური ანოტირება, გაანალიზებულია მორფოლოგიური ანოტირებისას გამოვლენილი შეცდომები.
11. სადისერტაციო ნაშრომში ასევე წარმოვადგინეთ ქართული ენის სიხშირული სია (frequency list), ყველაზე ხშირად ხმარებული 17,000 სიტყვა-ფორმა (არა ლემა) ქართული ენის ვებკორპუსის მიხედვით. აღნიშნული სიხშირული სია განისაზღვრა კვლევის პირველ ეტაპზე, მორფოლოგიური ანოტირებისა და ლემატიზების განხორციელებამდე. ქართული ენის სიხშირული სია საინტერესოა იმ თვალსაზრისით, რომ იძლევა ინფორმაციას ყველაზე ხშირად ხმარებული სიტყვა-ფორმების შესახებ, რაც ეჭვგარეშეა, მნიშვნელოვანია ენის კვლევისათვის. გარდა ამისა, მოცემული სიხშირული სია ამ ნაშრომის ერთ-ერთ სიახლეს წარმოადგენს.

12. სადისერტაციო ნაშრომში განსახდვრულია ქართული ენის სიხშირული ლექსიკა და აღწერილია ამ მიზნით შესრულებული სამუშაოები. საუბარია ვებკორპუსის ლემაზიტებაზე. ჩვენს კვლევაში გამოყენებულია CST ლემატიზების ინსტრუმენტი (CST Lemmatiser), რომელიც შემუშავებულია კოპენჰაგენის უნივერსიტეტის ენის ტექნოლოგიის ცენტრში. ქართული ენის ვებკორპუსის ლემატიზება რამდენიმე ეტაპად წარიმართა:

- 1) ლემატიზების პირველ საფეხურზე კორპუსიდან ამოვიღეთ ყველაზე ხშირად ხმარებული 35,000 სიტყვა-ფორმა მორფოლოგიური ტაგებით;
- 2) მეორე ეტაპზე კი კორპუსიდან ამოღებული მეტყველების ნაწილების ტაგების (POS – Part-of-Speech Tags) მქონე 35,000 სიტყვა-ფორმა ხელით დავამუშავეთ - მიუწერეთ შესაბამისი ლემები. ხელით დამუშავებული ლემატიზებული მასალის საფუძველზე მოვამზადეთ საწვრთნელი მასალა (training data) კორპუსის ავტომატური ლემატიზებისათვის.
- 3) ლემატიზების მესამე ეტაპზე ხელით დამუშავებული 35,000-სიტყვიანი საწვრთნელი მასალისა და CST ლემატიზების ინსტრუმენტის საშუალებით მოვახდინეთ მთლიანი კორპუსის ავტომატური ლემატიზება.
- 4) ლემატიზების პროცესის მეოთხე ეტაპზე ამოვიღეთ ავტომატურად ლემატიზებული ყველაზე ხშირად ხმარებული ლემები კორპუსიდან, დაახლოებით 18,000 ლემა. მიღებული ლემები დავამუშავეთ ხელით, გავფილტრეთ და გამოვრიცხეთ ისეთი შემთხვევები, როდესაც ლემები არასწორად იყო შერჩეული სიტყვა-ფორმისათვის და შედეგად წარმოვადგინეთ 10,000 ყველაზე ხშირად ხმარებული სიტყვა.

13. სადისერტაციო ნაშრომის ძირითადი სიახლეს ქართული ენის სიხშირული ლექსიკა - 10,000 ყველაზე ხშირად ხმარებულ სიტყვა, რომელთაც მითითებული აქვს სიხშირული ინდექსი და ნაჩვენებია ამა თუ იმ კონკრეტული სიტყვის სიხშირული განაწილება (frequency distribution) კორპუსში. გარდა ამისა, აღნიშნული სიტყვიანი გაწყობილია ანბანთრიგზე.

14. ქართული ენის სიხშირული ლექსიკის შერჩევისას ლემების წარმოდგენის პრინციპებად გამოვიყენეთ შემდეგი: სახელები მოცემულია მხოლოდობითი რიცხვის სახელობითი ბრუნვის ფორმებით; ხოლო ზმნა წარმოდგენილია მესამე პირის მხოლოდობითი რიცხვის აწმყოსა და მყოფადის მწკრივებში.
15. სადისერტაციო ნაშრომს ელექტრონული სახით დართული აქვს ქართული ენის სიხშირული ლექსიკა – ყველაზე ხშირად ხმარებული 10 000 სიტყვის ორი ვერსია: პირველი ვერსია დალაგებულია სიხშირული ზრდის მიხედვით, ხოლო მეორე ვერსია კი ანაბანის მიხედვით და მითითებული აქვს შესაბამისი სიხშირული ინდექსი.
16. სადისერტაციო ნაშრომში წარმოდგენილია ქართული ენის სიხშირული „ენ-გრამები“, რომელთაც განვიხილავთ როგორც მომიჯნავე სიტყვების რეალიზაციას. ორკომპონენტიანი, სამკომპონენტიანი, ოთხკომპონენტიანი და ხუთკომპონენტიანი „ენ-გრამების“ სრული ცხრილი ნაშრომს ელექტრონულად აქვს დართული.
17. სადისერტაციო ნაშრომის ბოლო, მეოთხე თავში განხილულია ქართული ენის ვებკორპუსი - KaWaC, როგორც წყარო ქართულ-ვეროპული ტიპის სასწავლო ლექსიკონის სიტყვა-სტატიების შესადგენად. ამ მიზნით, ვებკორპუსის ბაზაზე შევადგინეთ სიტყვა-სტატიები შევადარეთ ქართული ენის განმარტებითი 8-ტომეული ლექსიკონისა და გიორგი ცოცანიძის, ნანა ლოლაძის და ქეთევან დათუკიშვილის „ქართულ ლექსიკონის“ შესაბამის სიტყვა-სტატიებს. როგორც შედარებამ აჩვენა, ქართული ენის ვებკორპუსის მიხედვით შედგენილი სიტყვა-სტატიები ბევრად ინფორმაციულია და მოიცავს ისეთ ახალ მნიშვნელობებსა და რეალიებს, რაც ქართული ენის ლექსიკონებში არ არის ასახული და ვფიქრობთ, რომ მათი ასახვა სასწავლო ტიპის ლექსიკონებში აუცილებელია. ვებკორპუსის ბაზაზე შედგენილი სიტყვა-სტატიები ასევე მდიდარია თანამედროვე ქართულ ენაში გავრცელებული შესიტყვებებითა და იდიომებით. ჩატარებულმა ანალიზმა ცხადყო, რომ ჩვენ მიერ შექმნილი ქართული ენის ვებკორპუსი მეტად ეფექტურია ქართული სალექსიკონო სიტყვა-სტატიების შესადგენად, სიტყვების ყველაზე ხშირად ხმარებული



პოლისემიური მნიშვნელობების, შესიტყვებებისა და იდიომების გამოსავლენად.

18. ნაშრომში წარმოდგენილი უახლესი კორპუსული მეთოდები და მიდგომები შეიძლება გამოყენებულ იქნეს სხვა ქართველური ენების კორპუსების, კორპუსული რესურსების შექმნისათვის და მსგავსი კორპუსული კვლევის განსახორციელებლად.

## გამოყენებული ლიტერატურა

1. ბერიძე, სურმავა 2009: ბერიძე მ., სურმავა ნ., ქართული დიალექტური კორპუსი – როგორც ლინგვისტური საკვლევი ინსტრუმენტი, საერთაშორისო კონფერენცია „ჰუმანიტარული მეცნიერებები ინფორმაციულ საზოგადოებაში, ბათუმი, 2009 (<http://server.itex.ge/GDC>).
2. გოგოლაშვილი 2011: გოგოლაშვილი გ. თანამედროვე ქართული ენის მორფოლოგია, თბილისი. 2011.
3. დათუკიშვილი, ლოლაძე, ცოცანიძე 2014: დათუკიშვილი ქ., ლოლაძე ნ., ცოცანიძე გ. „ქართული ლექსიკონი“, ბაკურ სულაკაურის გამომცემლობა. თბილისი, 2014
4. მარგალიტაძე, ქერეჭაშვილი 2013: მარგალიტაძე თ., ქერეჭაშვილი გ. სამეცნიერო ტექსტების ინგლისურ-ქართული პარალელური კორპუსი. „ქართული ენა და თანამედროვე ტექნოლოგიები-2013“, კონფერენციის მასალები. თბილისი, 2013
5. შანიძე 1973: შანიძე ა. ქართული ენის გრამატიკის საფუძვლები: თხზულებანი. ტ. 3, ნაწ. 1. მორფოლოგია: მეორე გამოცემა, 1973
6. ჩიქობავა 1952: ჩიქობავა ა. ენათმეცნიერების შესავალი, თბილისი, 1952
7. ჩიქობავა 1964: ჩიქობავა ა. ქართული ენის განმარტებითი ლექსიკონი
8. Baayen 2001: Baayen, A. Word frequency distributions. Dordrecht: Kluwer. 2001.
9. Baroni & Bernardini 2004: Baroni, M., and Bernardini S. Bootcat: Bootstrapping corpora and terms from the web. In Proceedings of LREC2004. Lisbon. 2004.
10. Baroni & Bernardini 2006: Baroni, M. & Bernardini, S. Wacky! Working papers on the Web as Corpus. Bologna: Gedit. 2006.
11. Baroni & Kilgarriff 2006: Baroni, M., and Kilgarriff A. Large linguistically-processed Web corpora for multiple languages. In Companion Volume to Proceedings of the European Association of Computational Linguistics, 87–90. Trento. 2006.
12. Baroni & Ueyama 2006: Baroni, M. & Ueyama, M. Building general- and special-purpose corpora by web crawling. In: Proceedings of the 13th NIJL International Symposium, Language Corpora: Their Compilation and Application. Tokyo, Japan, pp. 31-40.2006.

13. Baroni, Bernardini, Ferraresi, Zanchetta 2008: Baroni M., Bernardini S., Ferraresi A., Zanchetta E. *The WaCky Wide Web: A Collection of Very Large Linguistically Processed Web-Crawled Corpora*. 2008.
14. Baroni, Chantree, Kilgarriff, Sharoff 2008: Baroni, M., F. Chantree, A. Kilgarriff, and S. Sharoff *CleanEval: A competition for cleaning web pages*. In *Proceedings of the 6th Language Resources and Evaluation Conference, LREC 2008*. Marrakech. 2008.
15. Belz & Kilgarriff 2006: Belz, A. & A. Kilgarriff. *Shared-task evaluations in HLT: Lessons for NLG*. *Proceedings of the 4th International Conference on Natural Language Generation (INLG'06)*, 133-135.
16. Biber, Johansson, Leech, Conrad, Finegan 1999: D. Biber, S. Johansson, G. Leech, S. Conrad, and E. Finegan *Longman grammar of spoken and written English*. Harlow, London.
17. Biemann, Giuliano, & Gliozzo 2007: Biemann, C., Giuliano, C., & Gliozzo A. *Unsupervised part-of-speech tagging supporting supervised methods*. In *Proceedings of recent advances in natural language processing (RANLP-07)*, Borovets, Bulgaria. 2007.
18. Brants 2000: Brants, T. *TnT: A statistical part-of-speech tagger*. In *Proceedings of the sixth conference on applied natural language processing (ANLP-00)* (pp. 224–231). San Francisco, CA, USA: Morgan Kaufmann Publishers Inc.
19. Brill 1992: Brill, E. *A simple rule-based part of speech tagger*. In *Proceedings of the third conference on applied natural language processing (ANLP-92)* (pp. 152–155). Morristown, NJ, USA: Association for Computational Linguistics.
20. Broder, Glassman, Manasse, Zweig 1997: Broder, A., Glassman, S., Manasse, M. & Zweig, G. *Syntactic clustering of the web*. In: *Proceedings of the Sixth International World WideWeb Conference*. Santa Clara, California, pp. 391-404.
21. Brown 1996: Brown, N. J. *Russian learners' dictionary: 10,000 words in frequency order*. Routledge, London. 1996.
22. Cermák & Kren 2010: Cermák, F. and Kren, M. *A frequency dictionary of Czech: core vocabulary for learners*. Routledge. 2010.
23. Charniak, Hendrickson, Jacobson, Perkowski 1993: Charniak, E., Hendrickson, C., Jacobson N., & Perkowski, M. *Equations for part-of-speech tagging*. In *National Conference on Artificial Intelligence* (pp. 784–789).
24. Cheng, Greaves, Warren 2006: Cheng, W., Greaves, C., and Warren, M. *From n-gram to skipgram to concgram*. *International Journal of Corpus Linguistics*, 11(4): 411–33.

25. Chowdhury, Frieder, Grossman, McCabe 2002: Chowdhury, A., Frieder, O., Grossman, D., and McCabe, M. Collection statistics for fast duplicate document detection. *ACM Transactions on Information Systems*, 20(2): 171–91.
26. Ciaramita & Baroni 2006: Ciaramita, M. & Baroni, M. Measuring web corpus randomness: A progress report. In: Baroni and Bernardini (2006), pp. 127-158. 2006.
27. Clark 2000: Clark, A. Inducing syntactic categories by context distribution clustering. In Cardie, C., Daelemans, W., Nédellec, C., & Tjong Kim Sang, E., (Eds.), In *Proceedings of the fourth conference on computational natural language learning and of the second learning language in logic workshop*, Lisbon, 2000 (pp. 91–94). Somerset, New Jersey: Association for Computational Linguistics.
28. Clark 2003: Clark, A. Combining distributional and morphological information for part of speech induction. In *Proceedings of the tenth conference on European chapter of the Association for Computational Linguistics (EACL-03)* (pp. 59–66). Morristown, NJ, USA: Association for Computational Linguistics.
29. Daraselia & Sharoff 2014: Daraselia S., Sharroff S. Morphosyntactic Specifications for KaWaC, a Web Corpus for Georgian, humanitaruli mecnierebebi informaciuli sazogadoebaSi II, baTumis SoTa rusTavelis saxelobis saxelmwifo universiteti, baTumi, saqarTvelo, 2014.
30. Daraselia et al: 2014: Daraselia at al. Towards Creating a Large Corpus of Georgian, 7th BIENNIAL IVACS CONFERENCE, Newcastle, UK, 2014
31. Davies & Gardner 2010: Davies, M. and Gardner, D. *A Frequency Dictionary of Contemporary American English*. Routledge. 2010.
32. Dickinson & Meurers 2003: Dickinson, M., Meurers, W.D. Detecting errors in part-of-speech annotation. In: *Proceedings of the 10th Conference of the European Chapter of the Association for Computational Linguistics (EACL-03)*.
33. Elworthy 1995: Elworthy D. Tagset design and inflected languages. In *7th Conference of the European Chapter of the Association for Computational Linguistics (EACL), From Texts to Tags: Issues in Multilingual Language Analysis SIGDAT Workshop*, pages 1–10, Dublin. 1995.
34. Erjavec 2004: Erjavec T. MULTEXT-East Version 3: Multilingual Morphosyntactic Specifications, Lexicons and Corpora. In *Fourth International Conference on Language Resources and Evaluation, LREC'04*, pages 1535 – 1538. 2004.

35. Ertöz, Steinbach, Kumar 2002: Ertöz, L., Steinbach, M., & Kumar, V. A new shared nearest neighbor clustering algorithm and its applications. In Proceedings of workshop on clustering high dimensional data and its applications (pp. 105–115).
36. Fairon, Naets, Kilgarriff, Schryver 2007: Fairon C., Naets H., Kilgarriff A., G.-M. de Schryver. Building and exploring web corpora – Proceedings of the 3rd Web as Corpus Workshop, incorporating Cleaneval. Presses Universitaires de Louvain, Louvain. 2007
37. Feldman, Hana, Brew 2006: Feldman, A., Hana, J., and Brew, Ch. A cross-language approach to rapid creation of new morpho-syntactically annotated resources. In Proc. LREC 2006.
38. Ferraresi 2007: Ferraresi A. Building a very large corpus of English obtained by web crawling: ukWaC. Master's thesis, University of Bologna. Retrieved January 28, 2008 from <http://wacky.sslmit.unibo.it>. 2007.
39. Finch & Chater 1992: Finch, S., & Chater, N. Bootstrapping syntactic categories using statistical methods. In Background and experiments in machine learning of natural language: Proceedings of the 1st SHOE Workshop (pp. 229–235). Brabant, Holland: Katholieke Universiteit.
40. Fletcher 2004: W. Fletcher. Making the web more useful as a source for linguistic corpora. In U. Connor and T. Upton, editors, Corpus Linguistics in North America 2002, Amsterdam. Rodopi.
41. Forsyth & Sharoff 2011: Forsyth, R. and Sharoff, S. From crawled collections to comparable corpora: An approach based on automatic archetype identification. In Proc. Corpus Linguistics Conference, Birmingham. 2011.
42. Freitag 2004a: Freitag, D. Toward unsupervised whole-corpus tagging. In Proceedings of the 20th international conference on Computational Linguistics (COLING-04) (p. 357). Morristown, NJ, USA: Association for Computational Linguistics.
43. Freitag 2004b: Freitag, D. Trained named entity recognition using distributional clusters. In Proceedings of EMNLP-04, (pp. 262–269).
44. Giménez & Márquez 2004: Giménez, J. & Márquez, L. Svmtool: A general pos tagger generator based on support vector machines. In Proceedings of the Forth Language Resources and Evaluation Conference, LREC 2004, Lisbon.
45. Hajic & Hladká 1998: Hajic, J. & Hladká, B. Tagging inflective languages: Prediction of morphological categories for a rich, structured tagset. In Proceedings of COLING-ACL, pages 483–490.  
<http://www.ncl.ac.uk/ecls/news/conferences/IVACS2014/>

46. Hundt, Nesselhauf & Biewer 2007: Hundt, M., Nesselhauf, N. & Biewer, C. *Corpus linguistics and the web*. Amsterdam: Rodopi. 2007.
47. Joho & Sanderson 2004: Joho, H., and M. Sanderson. *The SPIRIT collection: An overview of a large web collection*.
48. Kessler, Nunberg, Schütze 1997: Automatic detection of text genre. In *Proceedings of the 35th ACL/8th EACL*, 32–38.
49. Kilgarriff & Baroni 2006: Kilgarriff, A. & Baroni, M. *Proceedings of the 2nd International Workshop on the Web as Corpus*. East Stroudsburg (PA): ACL. 2006.
50. Kilgarriff & Grefenstette 2003: Kilgarriff, A., Grefenstette, G. Introduction to the special issue on the Web as Corpus. *Computational Linguistics*, 29(3), 333–347. 2003.
51. Kilgarriff 1997: Kilgarriff A. Putting frequencies in the dictionary. *International Journal of Lexicography*, 10(2):135– 155. 1997
52. Kilgarriff 2001: Kilgarriff, A. The web as corpus. In *proceeding of corpus linguistics 2001*. Lancaster. URL <http://www.itri.bton.ac.uk/techreports/ITRI-01-14.abs.html>
53. Kluwer, Baroni, Bernardini 2006: Kluwer, Dordrecht. M. Baroni and S. Bernardini, *Wacky! Working papers on the Web as Corpus*, Bologna. Gedit.
54. Kornai, Halacsy, Nagy, Oravecz, Tron, , & Varga 2006: Kornai, A., Halacsy, P., Nagy, V. Oravecz, C., Tron, V., & Varga, D. Web-based frequency dictionaries for medium density languages. In: Kilgarriff and Baroni (2006), pp. 1-8. 2006.
55. Manning 2010: Christopher D. Manning. *Part-of-Speech Tagging from 97% to 100%: Is It Time for Some Linguistics?* Departments of Linguistics and Computer Science Stanford University. 2010.
56. Mehler & Gleim 2006: Mehler, A., and R. Gleim. The net for the graphs – towards webgenre representation for corpus linguistic studies. In *WaCky! Working papers on the Web as Corpus*, eds. M. Baroni and S. Bernardini. Bologna: Gedit. 2006.
57. Meyer, Stein 2004: Meyer zu Eissen, S., and B. Stein. Genre classification of web pages. In *Proceedings of the 27th German Conference on Artificial Intelligence*. Ulm.
58. Rehm, Santini, Mehler, Braslavski, Gleim, Stubbe, Symonenko, Tavosanis, Vidulin 2008: Rehm, G., M. Santini, A. Mehler, P. Braslavski, R. Gleim, A. Stubbe, S. Symonenko, M. Tavosanis, and V. Vidulin. Towards a reference corpus of web genres for the evaluation of genre identification systems. In *Proceedings of the 6th Language Resources and Evaluation Conference, LREC 2008*. Marrakech.
59. Santini & Sharoff 2007: Santini M. & Sharoff S. *Proceedings of the CL 2007 Colloquium: Towards a Reference Corpus of Web Genres*, Birmingham, UK. 2007

60. Santini 2007: Santini M. Characterizing genres of web pages: genre hybridism and individualization. In Proceedings of the 40th Hawaii International Conference on System Sciences, poster session, pages 1–10, Waikoloa, Hawaii. 2007.
61. Santini 2007: Santini, M. Automatic identification of genre in Web pages. Ph.D. thesis, University of Brighton. 2007.
62. Schmid 1994: Schmid, H. Probabilistic part-of-speech tagging using decision trees. In International Conference on New Methods in Language Processing, Manchester.
63. Schütze 1995: Schütze, H. Distributional part-of-speech tagging. In Proceedings of the 7th conference on European chapter of the association for Computational Linguistics (EACL-95) (pp. 141–148), San Francisco, CA, USA: Morgan Kaufmann Publishers Inc. 1995.
64. Segalovich 2003: Segalovich, I. A fast morphological algorithm with unknown word guessing induced by a dictionary for a web search engine. In Proc. of MLMTA-2003, Las Vegas.
65. Sharoff 2005: Sharoff S. Methods and tools for development of the Russian Reference Corpus. In Archer, D., Wilson, A., and Rayson, P., editors, Corpus Linguistics Around the World, pages 167–180. 2005.
66. Sharoff 2006: Sharoff S. Creating general-purpose corpora using automated search engine queries. In M. Baroni and S. Bernardini, editors, Wacky! Working papers on the Web as Corpus, pages 63–98, Bologna. Gedit. 2006
67. Sharoff 2006: Sharoff, S. Creating general-purpose corpora using automated search engine queries. In: Baroni and Bernardini (2006), pp. 63-98. 2006
68. Sharoff 2006: Sharoff, S. Open-source corpora: using the net to fish for linguistic data. International Journal of Corpus Linguistics, 11(4):435–462. 2006.
69. Sharoff 2007: Sharoff S. Classifying web corpora into domain and genre using automatic feature identification. In Proceedings of Web as Corpus Workshop. Louvain-la-Neuve. 2007.
70. Sharoff 2007: Sharoff, S. Classifying web corpora into domain and genre using automatic feature identification. In Proc. of Web as Corpus Workshop, Louvain-la-Neuve. 2007.
71. Sharoff 2010: Sharoff S. In the Garden and in the Jungle Comparing Genres in the BNC and Internet. 2010.
72. Sharoff, Kopotev, Erjavec, Feldman, Divjak 2008: Sharoff, S., Kopotev, M., Erjavec, T., Feldman, A., and Divjak, D. Designing and evaluating a Russian tagset. In Proceedings

- of the Sixth Language Resources and Evaluation Conference, LREC 2008, Marrakech. 2008.
73. Sharoff, S. (2005). Methods and tools for development of the Russian Reference Corpus. In Archer, D., Wilson, A., and Rayson, P., editors, *Corpus Linguistics Around the World*, pages 167–180. Rodopi, Amsterdam.
  74. Thelwall 2005: Thelwall M. Creating and using web corpora. *International Journal of Corpus Linguistics*, 10(4):517–541. 2005.
  75. Xiao, Rayson, McEnery 2009: Xiao, R., Rayson, P., and McEnery A frequency dictionary of Mandarin Chinese: core vocabulary for learners. Routledge. 2009
  76. Yimam, Gurevych, Eckart, Biemann 2013: Yimam, S.M., Gurevych, I., Eckart de Castilho, R., and Biemann C. WebAnno: A Flexible, Web-based and Visually Supported System for Distributed Annotations. In *Proceedings of ACL-2013, demo session*, Sofia, Bulgaria.
  77. Yimam, Eckart, Gurevych, Biemann 2014: Yimam, S.M., Eckart de Castilho, R., Gurevych, I., Biemann C. Automatic Annotation Suggestions and Custom Annotation Layers in WebAnno In: *Proceedings of ACL-2014, demo session*, Baltimore, MD, USA. 2014.



გამოყენებული ვებგვერდები:

1. <http://www.oxfordlearnersdictionaries.com/definition/english/>
2. <http://global.longmandictionaries.com/>
3. <http://www.collins.co.uk/category/English+Language+Teaching/COBUILD+Reference>
4. <http://dictionary.cambridge.org/dictionary/british/>
5. <http://www.macmillandictionary.com/>
6. <http://clu.uni.no/icame/brown/bcm.html>
7. <http://www.helsinki.fi/varieng/CoRD/corpora/HelsinkiCorpus/>
8. <http://clu.uni.no/icame/manuals/WELLMAN/INDEX.HTM>
9. <http://clu.uni.no/icame/manuals/LONDLUND/INDEX.HTM>
10. <http://www.coventry.ac.uk/bawe>
11. <http://linguistica.sns.it/CoLFIS/Home.htm>
12. <http://armazi.uni-frankfurt.de/framee.htm>
13. <http://clarino.uib.no/gekko/corpus-list>
14. <http://armazi.uni-frankfurt.de/gnc/gnc.htm>
15. <http://www.catedratelefonica.upf.es/cucweb>
16. <https://www.sketchengine.co.uk/documentation/wiki/Corpora/UKWaC>
17. <http://maltparser.org/>
18. <http://corpus.leeds.ac.uk/internet.html>
19. <http://nl.ijs.si/ME/V4/msd/html/index.html>
20. <http://www.collins.co.uk/category/English+Language+Teaching/COBUILD+Reference/>
21. <http://webdante.com/>
22. <https://skell.sketchengine.co.uk/run.cgi/skell>.