

# ქართული ენის ეროვნული კორპუსი - მნიშვნელობა და პერსპექტივები

პროფ. მანანა თანდაშვილი  
ფრანკფურტის უნივერსიტეტი  
e-Mail: tandaschwili@em.uni-frankfurt.de

## I. რა არის ენის ეროვნული კორპუსი

ცნება „კორპუსი“ (ან უფრო ზუსტად - „ტექსტური კორპუსი“ **Textcorpus**) საზოგადოდ აღნიშნავს რომელიმე მოცემული ენისათვის არსებული რესურსების ნაკრებს, რომელიც, როგორც წესი, ელექტრონულ მატარებლებზე არის გადატანილი და მოიცავს არა მარტო წერილობითი ტექსტების, არამედ წერილობით დოკუმენტირებული ზეპირმეტყველების „გაციფრებულ“ მასალებს. ამდენად, იგი აერთიანებს ენის რეალიზაციის ორივე ფორმას: ე.წ. „დაგეგმილ“ და „დაუგეგმავ“ ენას.

ენის კორპუსი შეიძლება განვიხილოთ, როგორც ენის ან მისი რომელიმე ქვესისტემის ვირტუალური მიკრომოდელი (ანუ ენის ელექტრონული რესურსების საწყობი) - დიდი მოცულობის ელექტრონული ტექსტური მასივი, რომელიც აღჭურვილია საგანგებო საკვლევითი ინსტრუმენტით ე. წ. „კორპუსის მენეჯერით.“

ენის კორპუსები შეიძლება იყოს დალაგებული როგორც თემატურად (მაგ. ქართული წარწერების კორპუსი, ქართული ხალხური ზღაპრების კორპუსი, ქართული იდიომატური გამოთქმების კორპუსი, ქართული ლიტერატურის კორპუსი, პრესის კორპუსი...), ასევე ქრონოლოგიურად, რომელიმე ეტაპზე შეზღუდული. მაგ. XIX საუკუნის პრესის კორპუსი, საბჭოთა პერიოდის პრესის კორპუსი, ძველი, ახალი ან უახლესი ქართული ლიტერატურის კორპუსი.

ენის ეროვნული კორპუსი, ენის კორპუსისაგან განსხვავებით, წარმოადგენს სახელმწიფო ენისათვის შექმნილ ვირტუალურ სივრცეს, რომელშიც განთავსებულია სახელმწიფო ენის ისტორიული და თანამედროვე მემკვიდრეობა და 21-ე საუკუნის სახელმწიფოსათვის ისეთსავე აუცილებელ კომპონენტს წარმოადგენს, როგორც დამწერლობა, ლექსიკონი, ენციკლოპედია... და განიხილება სახელმწიფო ატრიბუტიკის - გერბის, დროშისა და ჰიმნის გვერდით.

ენის კორპუსსა და ენის ეროვნულ კორპუსს ერთადერთი საერთო ნიშანი აქვთ: ორივე კორპუსი მოცემული ენის სისტემის ამსახველ ვირტუალურ სივრცეს წარმოადგენს. განსხვავება მხოლოდ ისაა, რა მიზნით იქმნება იგი, რა აქტუალობა გააჩნია მას და ვინ არის მისი მომხმარებელი.

თუ ქართული ენის მაგალითზე განვიხილავთ, ეს შედარება ასე გამოიყურება:

1. ქართული ენის კორპუსის ფორმა არის ციფრული (დიგიტალური) და წარმოადგენს ქართული ენის სისტემის ამსახველ ვირტუალურ სივრცეს;
2. მისი მიზანია ენის ელექტრონული დოკუმენტაცია, ენის კვლევა თანამედროვე ტექნოლოგიების გამოყენებით განხორციელდეს;
3. მისი აქტუალობა მდგომარეობს კვლევის ეფექტურობაში (საკვლევი ველის მოცულობა და მიების სისწრაფე) და გამოირჩევა ვერიფიკაციის მაღალი ხარისხით (სულ რამდენიმეწუთიანი ოპერაციის შედეგად შესაძლებელია ნებისმიერი თეორიული პოსტულატის ვერიფიკაცია ან ფალსიფიკაცია).

განსაკუთრებული მნიშვნელობა ენიჭება ენის კორპუსების შექმნა, ანუ ენის ელექტრონულ დოკუმენტაციას, გაქრობის საშიშროების წინაშე მდგარი ენების შემთხვევაში ენის რევიტალიზაციის თვალსაზრისით.

ქართული ენის ეროვნული კორპუსის შემთხვევაში, განსხვავებულია კორპუსის შექმნის მიზანი და აქტუალობა.

1. ქართული ენის ეროვნული კორპუსი წარმოადგენს სახელმწიფო ენის სისტემის ამსახველ ვირტუალურ სივრცეს.

2. ეროვნული კორპუსის შექმნის მიზანია:

- ეფექტური ენობრივი პოლიტიკის გატარება;
- ქვეყნის საგანმანათლებლო სისტემის მოდერნიზაცია;
- კომპიუტერული ტექნოლოგიების შექმნა სახელმწიფო ენისათვის და მისი დანერგვა საზოგადოებრივ ცხოვრებასა და სამეცნიერო დარგებში;
- საინფორმაციო სივრცეში სახელმწიფო ენის სრულფასოვანი ფუნქციონირების უზრუნველყოფა.

3. ენის ეროვნული კორპუსის აქტუალობა გამოიხატება იმით, რომ იგი უზრუნველყოფს ქართული ენის, როგორც სახელმწიფო ენის, სრულფასოვან და ეფექტურ ფუნქციონირებას სახელმწიფო ინტერესების ფარგლებში.

ჩვენ მიერ წარმოდგენილი პოსტულატების ადეკვატურობის შესამოწმებლად შეგვიძლია ინგლისური ენის მაგალითი მოვიტანოთ: მიუხედავად იმისა, რომ ამერიკის, ბრიტანეთისა და ავსტრალიის სახელმწიფო ენა ინგლისურია, ერთმანეთისაგან დამოუკიდებლად არსებობს ამერიკის ეროვნული კორპუსი, ბრიტანეთის ეროვნული კორპუსი და ავსტრალიის ეროვნული კორპუსი და ახალი ზელანდია, რომლის სახელმწიფო ენას ადგილობრივი, პოლინეზიური მათრი-ენის გვერდით ინგლისური ენა წარმოადგენს, ვერცერთ აქ აღნიშნულ კორპუსს ვერ გამოიყენებს სახელმწიფო ინტერესების შესაბამისად, ვინაიდან ნებისმიერი ეროვნული კორპუსი მხოლოდ მოცემული ქვეყნისათვის შეიძლება იყოს სახელმწიფო ენის სრულფასოვანი ფუნქციონირების გარანტი და მოქნილი იარაღი.

## II. რა სახელმწიფოებრივი გამოყენება შეიძლება ჰქონდეს ქართული ენის ეროვნულ კორპუსს?

ქართული ენის ეროვნული კორპუსის უპირველეს პრიორიტეტად სახელმწიფოებრივი გამოყენების თვალსაზრისით ენობრივი პოლიტიკის გატარება მიგვაჩნია, რაც შემდეგში გამოიხატება:

- ქართული ენის **ნორმალიზაცია** და **სტანდარტიზაცია** - აქ, უპირველეს ყოვლისა, იგულისხმება ეროვნული კორპუსის მნიშვნელობა სახელმწიფო ენის ნორმების დამდგენი კომისიისათვის, გნებავთ ენის პალატისათვის ან სხვა სახის სახელმწიფო უწყებისათვის ენობრივი ნორმების სტანდარტიზაციის მიზნით;
- ქართული ენის **სრულფასოვანი ფუნქციონირების უზრუნველყოფა**. აქ იგულისხმება როგორც ტექნიკური (სპელჩეკერისა და პარსერის დახვეწა ქართული ენისათვის), ისე იურიდიული ასპექტი, როგორცაა, მაგ. ქართული ენის, როგორც სახელმწიფო ენის სავალდებულო სწავლება ქვეყნის მასშტაბით, მათ შორის არაქართველი მოსახლეობისთვისაც.

აქვე გვსურს, ხაზი გავუსვათ ერთ გარემოებასაც, რომელიც განსაკუთრებით დიასპორების სამინისტროსთვის შეიძლება იყოს საინტერესო: ეს გახლავთ ენობრივი პოლიტიკის სახელმწიფო ინტერესების გატარება ქვეყნის ფარგლებს გარეთ - ქართული ენის გავრცელების ვირტუალური გეოგრაფიის გაფართოება - მისი ხელმისაწვდომობა როგორც ისტორიული დიასპორებისათვის (ირანში, აზერბაიჯანსა და თურქეთში), აგრეთვე თანამედროვე დიასპორული ჯგუფებისათვის მსოფლიოს ნებისმიერ ქვეყანაში;

უდიდესია ეროვნული კორპუსის მნიშვნელობა ქვეყნის საგანმანათლებლო სისტემის მოდერნიზაციის თვალსაზრისით. დღეს შეუძლებელია სწავლების თანამედროვე

მეთოდების შემოტანა და დანერგვა როგორც საშუალო, ისე უმაღლეს საგანმანათლებლო სისტემაში ელექტრონული რესურსების გარეშე, განსაკუთრებით - ელექტრონული სწავლების ფონზე. მიუხედავად იმისა, რომ საქართველოში უკვე იკიდებს ფეხს სწავლების ეს ფორმა, აუცილებლად და გადაუდებლად მიგვაჩნია eLearning-ის გაქართულება - ქართული ენის ადაპტაცია ელექტრონული სწავლების პლატფორმებისათვის.

ალბათ დამსწრე საზოგადოებას არც სჭირდება იმის მტკიცება, თუ რამდენად დიდია ეროვნული კორპუსის მნიშვნელობა ლექსიკოლოგიისათვის. აქ გამომსვლელი არაერთი მომხსენებელი უკვე შეეხო ლექსიკოლოგიის საკითხს და მე სიტყვას აღარ გავაგრძელებ. ელექტრონული ლექსიკონების (იქნება ეს განმარტებითი, ორთოგრაფიული, სინონიმთა თუ ორენოვანი ლექსიკონები) არსებობის გარეშე ალბათ შეუძლებელია 21 საუკუნის ქართველი ახალგაზრდების ინტელექტუალურ განვითარებაში ეფიციენტობის მიღწევა.

### III. ქართული ენის ეროვნული კორპუსის მეცნიერული გამოყენების სფეროები

ეროვნული კორპუსის არსებობა ერთ-ერთი წინაპირობაა საქართველოში ემპირიული ცნობიერების განვითარებისათვის, როგორც საზოგადოების ისე სამეცნიერო საზოგადოების ფარგლებში, განსაკუთრებით ეფექტური მეცნიერული კვლევის წარმართვის მიზნით.

კერძოდ: მეთოდოლოგიური თვალსაზრისით, იგი სერიოზულ ემპირიულ საფუძველს უქმნის ქართველოლოგიურ კვლევებს და ხელს უწყობს კორპუსზე ორიენტირებული კვლევების იმპლემენტაციას. რაც შეეხება პრაქტიკულ შედეგებს, ენის ეროვნული კორპუსი საშუალებას მოგვცემს ქართული ენის ისტორიის, სტრუქტურისა და განვითარების ტენდენციები თანამედროვე ტექნოლოგიური საშუალებებით ვიკვლიოთ.

გარდა ამისა, ეროვნული კორპუსის არსებობა საშუალებას მოგვცემს, სახელმწიფო ენისათვის უმნიშვნელოვანესი დარგი - **ლექსიკოგრაფია** სრულიად ახალ, თვისებრივად განსხვავებულ საფეხურზე ავიყვანოთ. აქ ჩვენ ვგულისხმობთ ტექსტიდან ლექსიკონების ელექტრონული ექსტრაქციების შესაძლებლობას.

### IV. ქართული ენის ეროვნული კორპუსის მომხმარებელი

მოგეხსენებათ, რომ დღეს ნებისმიერი პროდუქტის შექმნა და მისი ბაზარზე შეტანა მომხმარებლის პროგნოზირებად კოეფიციენტზე არის გათვლილი.

**ვინ შეიძლება იყოს ქართული ენის ეროვნული კორპუსის მომხმარებელი?**

ქართული ენისათვის აქამდე შექმნილი კორპუსების მოხმარების გამოცდილებას თუ გავითვალისწინებთ, ენის კორპუსის მომხმარებელი საზოგადოების ძალიან მცირე ნაწილია. ესენი არიან სამეცნიერო ფილოლოგიური საზოგადოების წარმომადგენლები (და ისიც შეზღუდულად), აგრეთვე ფილოლოგიური პროფილის სტუდენტთა ძალიან შეზღუდული რაოდენობა.

ეროვნული კორპუსის, უფრო სწორედ ეროვნულ კორპუსზე დაფუძნებული პროგრამების მომხმარებელი შეიძლება და უნდა გახდეს კიდევ ფართო საზოგადოება - განსხვავებული პროფესიისა და ინტერესების მქონე სხვადასხვა სოციალური და ასაკობრივი ჯგუფი.

აქვე გვინდა აღვნიშნოთ, რომ ეროვნული კორპუსების სამეცნიერო და სახელმწიფოებრივი გამოყენების დიდი გამოცდილება არსებობს მსოფლიოში, რომელიც აუცილებლად უნდა გავითვალისწინოთ. გარდა დიდი სახელმწიფოებისა, როგორებიცაა

ამერიკა, გერმანია, საფრანგეთი, ეროვნული კორპუსი შექმნილი აქვს და მის შესაქმნელად ძალისხმევა არ დაიშურეს, უპირველეს ყოვლისა, სახელმწიფოებრიობის მქონე მცირე ერებმა. შემთხვევით არ გვიხსენებია ჩვენი მოხსენების დასაწყისში ეროვნული კორპუსის, როგორც სახელმწიფო ატრიბუტიკის აუცილებელი შემადგენელი კომპონენტის მნიშვნელობა. იგი სახელმწიფოს, როგორც ცივილიზებული სამყაროს სრულფასოვანი წევრის, სავიზიტო ბარათს წარმოადგენს.

დასახელებული კორპუსების გარდა, დღეისათვის არსებობს რუსული, უნგრული, ჩეხური, ხორვატული, პოლონური, ბულგარული, ალბანური, ბერძნული, თურქული, სლოვაკური, (აღმოსავლეთ) სომხური ეროვნული კორპუსები... საგანგებოდ გვინდა გამოვყოთ სომხური ეროვნული კორპუსი. საჭირო ინტელექტუალური რესურსების არქონის ფონზე, სომხეთის მთავრობას დრო არ დაუკარგავს ეროვნული კადრების მომზადებასა და კვალიფიკაციის ამაღლებისათვის და ეროვნული კორპუსის შესაქმნელად რუსი მეცნიერები მოიწვია.

#### IV. ქართული ენის ეროვნული კორპუსი

დღევანდელი ჩემი მოხსენება არ გახლავთ პირველი მოხსენება ამ თემაზე და არც დღევანდელი შეკრება გახლავთ პირველი, სადაც მე და ჩემი კოლეგები მოგვიწვიეს ამ თემაზე სასაუბროდ. უკვე რამდენიმე წელია, ვცდილობთ ხელისუფლების დარწმუნებას იმაში, თუ რამდენად მნიშვნელოვანია ქართული სახელმწიფოსათვის ეროვნული კორპუსის შექმნა და რომ დღეს ქართული ენისათვის არსებული რესურსები სრულიად საკმარისია იმისათვის, რომ არ გადავდოთ საუკუნის მნიშვნელობის პროექტის დაწყება. სწორედ ამ პათოსით დაიწერა 2011 წელს პროექტი “ქართული ენის ეროვნული კორპუსი - ტექნოლოგიური ჩარჩოს მომზადება”, რომელიც ფოლკსვაგენის ფონდმა დააფინანსა და რომელიც, უამრავი დაბრკოლებების მიუხედავად, არსებობას განაგრძობს.

დღეისათვის სახეზეა ქართული ენის ეროვნული კორპუსის შესაქმნელად აუცილებელი ელექტრონული რესურსების დიდი ნაწილი: აღნიშნული რესურსები 30 წლის განმავლობაში იქმნებოდა როგორც უცხოეთის ქართველოლოგიური კვლევის ცენტრში - ფრანკფურტის უნივერსიტეტში (TITUS-ისა და ARMAZI-ს ელექტრონული ბაზები), ისე საქართველოს სამეცნიერო დაწესებულებებში - არნ. ჩიქობავას სახელობის ენათმეცნიერების ინსტიტუტში (ქართული ენის დიალექტური კორპუსი) და ივ. ჯავახიშვილის სახ. თბილისის სახელმწიფო უნივერსიტეტში. ამჟამად მიმდინარეობს ქართული ენის ეროვნული კორპუსისათვის ტექნოლოგიური ჩარჩოს მომზადება და არსებული რესურსების ტრანსპორტირება-ინტეგრაცია.

ქართული ენის ეროვნულ კორპუსს მიმდინარე ეტაპზე საფუძვლად დაედო ქართული ენისათვის შექმნილი ოთხი ელექტრონული რესურსი.

1. TITUS-ის ელექტრონული ტექსტების ბაზა (ფრანკფურტის უნივერსიტეტი)
2. ARMAZI -ს ელექტრონული ტექსტების ბაზა (ფრანკფურტის უნივერსიტეტი)
3. GEKKO - ქართული ენის ელექტრონული კორპუსის ანალიზატორი (ნორვეგია)
4. ქართული დიალექტური კორპუსი (საქართველო)

რაც შეეხება პროექტის ხელმძღვანელების გამოცდილებას: ფრანკფურტის უნივერსიტეტის ემპირიული ენათმეცნიერების ინსტიტუტს, რომელსაც გერმანელი ქართველოლოგი პროფ. იოსტ გიპერტი ხელმძღვანელობს, კორპუსების შექმნის დიდი გამოცდილება აქვს. გიპერტის იდეითა და უშუალო ხელმძღვანელობით დაიბადა TITUS-ი (ინდოევროპული ენების ტეზაურუსი), მისი კონცეფციითა და ჩვენი აქტიური მონაწილეობით შეიქმნა კავკასიურ ენათა ელექტრონული პლატფორმა ARMAZI. იოსტ

გიპერტი გახლავთ წამყვანი მეცნიერი ძველი გერმანულის რეფერენციული კორპუსისა Referenzkorpus Altdeutsch, რომელსაც ფრანკფურტის უნივერსიტეტი იენისა და ბერლინის ჰუმბოლდტის უნივერსიტეტთან თანამშრომლობით ახორციელებს. ორი წლის წინ კი ჰესენის მიწის ყველაზე დიდ, 5 მილიონი ევროს ღირებულების საერთაშორისო პროექტს **Digital Humanities – LOEWE** ჩაუდგა სათავეში, რომელსაც დარმშტადტის ტექნიკურ უნივერსიტეტთან და **ECoDA - European Cooperation on Digitizing Avestan** (ესპანეთი) თანამშრომლობით ახორციელებს.

ჩვენი სამეცნიერო ჯგუფი გახლდათ ერთ-ერთი პირველთაგანი, ვინც ნეიმეხენის მაქს პლანკის ინსტიტუტთან ერთად ზეპირმეტყველების ელექტრონული დოკუმენტაციის საგანგებო ფორმატი შეიმუშავა - ფაილის ახალი ტიპი, რომელიც საშუალებას იძლევა ვიდეო, აუდიო და ტექსტური მასალა ერთ ფორმატში განვითავსოთ.

ბერგენის უნივერსიტეტის პროფესორი **პაულ მოირერი** გახლავთ კომპუტერული ლინგვისტიკის სპეციალისტი, რომელსაც არაერთი კორპუსის შექმნაში აქვს მიღებული მონაწილეობა, მათ შორის ნორვეგიული ეროვნული კორპუსისა. იგი გახლავთ ავტორი ქართული კორპუსისა GEKKO, რომელსაც უკვე წლების განმავლობაში იყენებენ ქართული ენის სპეციალისტები. სწორედ პაულ მოირერის მიერ მუშავდება ეროვნული კორპუსის პროტოტიპული ვერსია, რომელიც აღჭურვილია მორფოლოგიური და სინტაქსური ანალიზატორით (პარსერით).

მეოთხე, ძალიან მნიშვნელოვანი რესურსი და, მე ვიტყვოდი, უპრეცედენტო შემთხვევა - ქართული ენის ეროვნული კორპუსის შესაქმნელად - გახლავთ არნ. ჩიქობავას სახ. ენათმეცნიერების ინსტიტუტში პროფ. მარინე ბერიძის მიერ შექმნილი “ქართული დიალექტური კორპუსი”. სამეცნიერო პროექტების შედეგად შექმნილ ამ კორპუსს, რომელსაც საქართველოს ეროვნული სამეცნიერო ფონდი აფინანსებდა, თავად მარინე ბერიძე წარმოგიდგინთ და მე აღარ გავაგრძელებ მასზე საუბარს.

### **რა აუცილებელი ტექნოლოგიური პრობლემები არის დღეისათვის გადაჭრილი კორპუსთან მიმართებით?**

უპირველეს ყოვლისა უნდა დავასახელოთ ქართული ანბანის საერთაშორისო კოდირების პრობლემა. არ შემიძლია არ აღვნიშნო, რომ ჯერ კიდევ გასული საუკუნის 90-იან წლებში, როდესაც საქართველოში ყველაფერი ინგრეოდა და ჩვენ თვალწინ ხდებოდა სახელმწიფო და საზოგადოებრივი დემონტაჟი, პარალელურად მიმდინარეობდა ქართული ანბანის საერთაშორისო კოდირებისათვის ზრუნვა. და ის ფაქტი, რომ უნიკოდში (კოდირების საერთაშორისო კონსორციუმი) დღეს არა მხოლოდ მხედრულის, არამედ ასომთავრულისა და ნუსხა-ხუცურის შრიფტი გაგვაჩნია. პროფ. გიპერტისა და ბატონ ირაკლი ღარიბაშვილის დამსახურება გახლავთ.

დამსწრე საზოგადოებას რომ არა მხოლოდ კორპუსის მოცულობაზე, არამედ მის შედგენილობაზე შეექმნას შთაბეჭდილება, მოკლედ აღვნიშნავთ, რომ დღეისათვის დიגיტალიზებულია თითქმის ყველა გამოცემული და ზოგიერთი გამოუცემელი ქართული ხელნაწერი, მათ შორის - უცხოეთის ბიბლიოთეკებსა და საცავებში დაცული პალიმფსესტური ხელნაწერებიც. მინდა ვისარგებლო შემთხვევით და მოგახსენოთ, რომ პალიმფსესტების დასამუშავებლად ჩვენმა ჯგუფმა პირველმა გამოიყენა მულტისპექტრალური აპარატი **MuSIS**-ი, რომელიც საქართველოს ეროვნულ მუზეუმს

გადაეცა საჩუქრად ექსპლოატაციისათვის, კერძოდ სვანური ხელნაწერების დიგიტალიზაციისთვის.

ამდენად, დღეისათვის ჩვენ გვაქვს არა მხოლოდ დიგიტალიზებული ტექსტები, არამედ შესაბამისი ფოტომასალაც, რაც ვერიფიკაციის უნიკალურ საშუალებას აძლევს მეცნიერებს ხელნაწერზე ვირტუალური მუშაობის დროს. თუმცა, აქვე გვინდა აღვნიშნოთ, რომ საქართველოში დაცული ხელნაწერების უმრავლესობა, მაგ. ხელნაწერთა ეროვნულ ცენტრში არსებული პალიმფსესტური ხელნაწერები (დაახლოებით 5000 ფურცელი), და სვანეთში (კერძო მფლობელობაში) დაცული ხელნაწერები, რომლებიც, სხვათა შორის, ძალიან მძიმე მდგომარეობაში იმყოფებიან (განსაკუთრებით ლახამულას ოთხთავი) გადაუდებელ რესტავრაციას და დიგიტალიზაციას საჭიროებენ.

## რა სტანდარტულ მოთხოვნებს უნდა აკმაყოფილებდეს კორპუსი? (კორპუსის პარამეტრები)

### 1. კორპუსის მოცულობა

უკვე არსებული ეროვნული კორპუსები სხვადასხვა მოცულობისაა: შდრ. ამერიკის ეროვნული კორპუსი - 22 მლნ ტოკენი, ბრიტანული ეროვნული კორპუსი - 100 მლნ. ტოკენი, სომხური ეროვნული კორპუსი - 110 მლნ. ტოკენი, რუსული ეროვნული კორპუსი - 150 მლნ. ტოკენი. დღეისათვის ჩვენს ხელთ არსებული რესურსების მოცულობა, რომელიც ქართულ ეროვნულ კორპუსს დაედება საფუძვლად, გახლავთ - **132 მლნ.** ტოკენი.

### 2. კორპუსის ბალანსირება

ვისაც კორპუსებთან შეხება ჰქონია, ის ალბათ დიდად ვერ მოიხიბლება მხოლოდ სტატისტიკური მონაცემებით. ყველას კარგად მოგვეხსენება, რომ ეროვნულ კორპუსზე მხოლოდ მაშინ შეიძლება საუბარი, როდესაც კორპუსი **ბალანსირებულია**. დღეისათვის ჩვენს ხელთ არსებული რესურსები ძირითადად ლიტერატურული ძეგლების ელექტრონული ვერსიებია. ეროვნული კორპუსი კი რეპრეზენტატიული უნდა იყოს არა მხოლოდ რაოდენობრივად, არამედ თვისებრივადაც - იგი უნდა წარმოგვიდგენდეს განსხვავებულ ჟანრულ რეპერტუარს ქართული ენისა: იქნება ეს ისტორიული დოკუმენტები, იურიდიული დოკუმენტები, სამეცნიერო ტექსტები, პრესის მასალები და ა.შ.

ვიდრე დავასრულებდე ჩემს გამოსვლას, მინდა ეროვნული კორპუსის სახელმწიფოებრივ მნიშვნელობაზე კიდევ ერთხელ მივაქცევინოთ ყურადღება დამსწრე საზოგადოებას და ერთი მაგალითი მოვიტანო გერმანული გამოცდილებიდან: ეს გახლავთ უახლესი ნასესხობებისა და ახალი კომპოზიციების ნორმალიზაციისა და სტანდარტიზაციის საკითხი გერმანულ სივრცეში.

უკვე ათწლეულზე მეტია (2000 წლიდან), რაც მანჰაიმის გერმანული ენის ინსტიტუტში IDS (Institut für deutsche Sprache) შეიქმნა გერმანული ენის “სიტყვის სადარაჯო” Die Wortwarte ([www.wortwarte.de](http://www.wortwarte.de)), რომლის იდეის ავტორი და განმახორციელებელი გახლდათ ლოთარ ლემნიცერი. ამ ინსტიტუტში ყოველდღიურად მიმდინარეობს გერმანული გაზეთების ელექტრონული ვერსიების „შემოწმება“ - მთელი იმ დღის მასალა პროგრამული უზრუნველყოფის შედეგად მიემართება გერმანულ 4 მილიარდსიტყვიან კორპუსს, ხდება ავტომატური ამოფილტვრა იმ დღეს პირველად გამოყენებული და „კორპუსისათვის უცნობი“ სიტყვებისა, რომლებიც

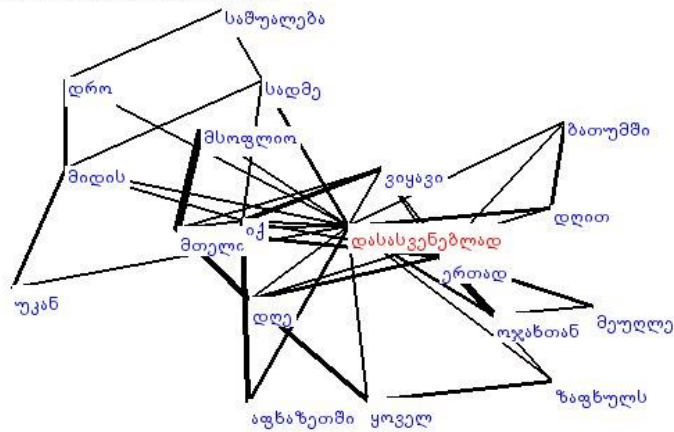
ნორმალიზაციისათვის გადაეგზავნება ლინგვისტთა სპეციალურ ჯგუფს; ეს ჯგუფი განიხილავს სიტყვის დაწერილობის სისწორეს და ადგენს/განმარტავს მის მნიშვნელობას. სიტყვის მოხმარების მიზანშეწონილებების შემთხვევაში, ხდება მისი სტანდარტიზაცია - ახალი სიტყვა თავის ადგილს პოულობს კორპუსში - ხდება შესაბამისი კონტექსტის კორპუსში იმპლემენტაცია. ასე, ყოველდღიურად იზრდება გერმანული ეროვნული კორპუსი ლინგვისტთა მეთვალყურეობის ქვეშ. ენისადმი სახელმწიფოებრივი დამოკიდებულების შესანიშნავი ნიმუშია თავად ლოთარ ლემნიცერის სიტყვები: „არსებობს ამინდის ბიურო (Wetterwarte), რომელიც აკვირდება ჰაერის მასების მოძრაობას და პროგნოზირებს ამინდს. არსებობს ობსერვატორია (Sternwarte), რომელიც აკვირდება ციური სხეულების შექმნას, მოძრაობასა და გაქრობას. ჩვენ შევქმენით „სიტყვის სადარაჯო“ (Wortwarte, აქ სიტყვათა თამაშია. შდრ. Wetterwarte და Sternwarte), რათა დავაკვირდეთ სიტყვების შექმნას, გერმანული ლექსიკის განვითარებას და ვამცნოთ იგი საზოგადოებას. ჩვენ დამკვირვებლები ვართ, ენის დამკვირვებლები. ჩვენი დაკვირვების ინტერპრეტაცია კი მათთვის მიგვინდია, ვისაც ეს ევალება“.

ცნობისათვის: გერმანიაში ყოველწლიურად ცხადდება კონკურსი „წლის სიტყვის“ გამოსავლენად.

და კიდევ ერთი მაგალითი გერმანული გამოცდილებიდან - ეს გახლავთ ინფორმაციის ვიზუალიზაცია და მისი გავლენა საზოგადოების ცნობიერების ჩამოყალიბებასა ან ცვლილებაზე. მოგეხსენებათ, რა დიდი გავლენა აქვს მედიას საზოგადოებრივ ცნობიერებაზე. დღეს უკვე აღარავინ არ დავობს იმის შესახებ, რომ მედიასაშუალებები საზოგადოებრივი ცნობიერების მანიპულირების ერთ-ერთ უმძლავრეს იარაღს წარმოადგენს. ინფორმაციის ვიზუალიზაციის თეორია („განწყობების წაკითხვა“) წარმატებით გამოიყენა ლაიფციგის უნივერსიტეტის პროფესორმა გერჰარდ ლაიერმა, რომლის პროგრამა 230 ენის, მათ შორის ქართული პრესის მასალებსაც (Opentext - ქართული პრესის ელექტრონული არქივი <http://www.opentext.org.ge> და Civil Georgia - ახალი ამბების სააგენტო სამენოვანი პორტალით <http://www.civil.ge>) მიემართება. პროგრამა ეფუძნება კოლოკაციისა (ტექსტში სიტყვის მეზობლად მდგარი მეორე სიტყვა) და კოკურენციის (ტექსტურ მასივში მოცემულ სიტყვასთან ყველაზე ხშირად გამოყენებული მეორე სიტყვა) ცნებას. მაგ. წინადადებაში **ზაფხულში დასასვენებლად ზღვაზე მივდივარ** სიტყვა **დასასვენებლად** კოლოკანტებია **ზაფხულში** და **ზღვაზე**. მაშინ როდესაც, ამ სიტყვის კოკურენტებად (ყველაზე ხშირად ხმარებულ სიტყვებად) ტექსტებში გვევლინება **ვიყავი** და **ოჯახთან ერთად**.



Graph v.1.6 für დასასვენებლად



სხვათა შორის, სიტყვა „დემონტაჟის“ კოოკურენტები „ძეგლისა“ და „შენობის“ გვერდით ამ ბაზაში სიტყვა „ხელისუფლებაც“ გახლავთ.

ძალიან ეფექტურია არა მხოლოდ ცალკეული სიტყვების, არამედ მთელი ციტატების ვიზუალიზაცია (კოოკურენციული კვლევა) ქრონოლოგიურ ჭრილში ეპოქების მიხედვით. მაგ. როდის აქტიურდებოდა ქართულ პრესაში (ანუ ქართველთა ცნობიერებაში) გამოთქმები „ქართლად ფრიადი ქუეყანად აღირაცხების...“ (გრიგოლ ხანძთელი) ან „...ვითარმედ ყოველი საიდუმლოდ ამას ენასა შინა დამარხულ არს... (იოანე ზოსიმე) და სხვა.

კორპუსის ამგვარი შესაძლებლობები კვლევის უდაოდ ეფექტურ საშუალებას უქმნის სოციოლოგებს, პოლიტოლოგებს, ისტორიკოსებსა და სხვა დარგის სპეციალისტებს. ინფორმაციის ვიზუალიზაცია (საზოგადოებრივ აზრის მაჯისცემის მოსმენა) არა მხოლოდ კვლევის, არამედ კონფლიქტების პროგნოზირების ეფექტური საშუალებაც გახლავთ. მეტიც, არსებობს კოოკურენტების „მართვის“ ონტოლოგია. ვინაიდან კოოკურენციები ძირითადად გარკვეული სიტყვების ერთდროული ხმარების სიხშირეზეა დამოკიდებული, იგი ადვილად მანიპულირებადია. მიზანმიმართული მცდელობით შესაძლებელია მათი სიხშირის გაზრდა (სასურველი კოოკურენტების ხშირი გამოყენება) ან ამა თუ იმ მოვლენისადმი პერსპექტივის შეცვლა.

ამის საუკეთესო მაგალითი გახლავთ ნიუ-იორკ თაიმსის ელექტრონული კორპუსიდან გამოთვლილი კოოკურენციების სურათი ცნებით „აბუ-გრაბი“ (Abu-Ghraib - ერაყში არსებული ციხე ამერიკული ჯარების მმართველობის ქვეშ). გამოკვლევის პირველ ქრონოლოგიურ ფაზაში ამ სიტყვის ირგვლივ დაჯგუფდა ისეთი ცნებები, როგორცაა „წამება“, „ადამიანის უფლებათა დარღვევა“, „არაჰუმანურობა“ და ასე შემდეგ. მეორე ქრონოლოგიურ ფაზაში „აბუ-გრაბ“- თან ასოცირებადი კოოკურენციები მოიცავდა ცნებებს „ამერიკის უსაფრთხოება“, „ტერორისტი“, „თერთმეტი სექტემბერი“ და ასე შემდეგ. ალბათ შეიძლება ვივარაუდოთ, რომ კოოკურენციის თემატური დისლოკაცია ამ სიტყვასთან (ფაქტთან) მიმართებაში ამერიკის ხელისუფლების მიზანმიმართული მონდომების შედეგი იყო.

და კვლავ რომ დავუბრუნდეთ ჩვენი მოხსენების ძირითად საკითხს, მოკლედ მიმოვიხილავთ იმ ინსტიტუციებს, რომლებიც წლების განმავლობაში მონაწილეობას



იღებდნენ ეროვნული კორპუსისათვის მოსამზადებელ სამუშაოში და ამჯერადაც ჩვენი პროექტის ქართველი პარტნიორები არიან:

- ივ. ჯავახიშვილის სახ. უნივერსიტეტი
- არნ. ჩიქობავას სახ. ენათმეცნიერების ინსტიტუტი
- ილიას სახელმწიფო უნივერსიტეტი
- ბათუმის შოთა რუსთაველის სახ. უნივერსიტეტი
- ქუთაისის აკაკი წერეთლის უნივერსიტეტი
- საქართველოს ეროვნული მუზეუმი
- საქართველოს ტექნიკური უნივერსიტეტი
- საქართველოს სამეცნიერო ბიბლიოთეკა

პროექტში ქართველ მეცნიერთა ყველა თაობაა ჩართული, უხუცესი მეცნიერებიდან დაწყებული - ჰუმანიტარულ მეცნიერებათა ფაკულტეტის სტუდენტების ჩათვლით. პროექტის მთლიანი ღირებულებაა 970 000 ევრო. პროექტის განხორციელების პირველ ეტაპზე ფოლკსვაგენის ფონდის მიერ გამოყოფილია 390 000 ევრო.

საგანგებოდ მინდა ხაზი გავუსვა იმ გარემოებას, რომ საქართველოში ბოლო წლებში „კორპუსული მოძრაობის“ საოცარი აღზევება შეინიშნება. ილიას უნივერსიტეტის მიერ განხორციელებული კორპუსული პროექტები, პარალელური კორპუსების სახით, ეროვნული კორპუსის ხვალისდელი დღეა. პროფ. თინა მარგალიტაძის „ორენოვანი ონლაინ ლექსიკონების“ პროექტს თავად ქალბატონი თინა წარმოადგენს, მაგრამ მინდა აქვე აღვნიშნო, რომ ასეთი ლექსიკონები ეროვნული კორპუსის შემდგომი განვითარების აუცილებელ წინაპირობას წარმოადგენს.

ჩემს გამოსვლას ამით დავასრულებ და დასკვნის სახით ვიტყვი, რომ ქართული ეროვნული კორპუსი ის აუცილებელი საძირკველია, რომელზედაც უნდა დაშენდეს ქართული ენის სახელმწიფოებრივ ფუნქციონირებასთან მოაზრებული ყველა რესურსი, სამეცნიერო კვლევებიდან დაწყებული, ენის ინოვაციური ტექნოლოგიებით დამთავრებული. ეროვნული კორპუსის შექმნა მხოლოდ დააჩქარებს გლობალურ ტექნოლოგიურ ქსელში საქართველოს ჩართვას.