ფრანკფურტის უნივერსიტეტის ემპირიული ენათმეცნიერების ინსტიტუტი

მანანა თანდაშვილი, მარიამ ყამარაული

შესავალი
დიგიტალურ ქართველოლოგიაში

ეძღვნება პროფესორ მზექალა შანიძის
95 წლის იუბილეს

თბილისი

2021 წელი

დიგიტალური ქართველოლოგია 21-ე საუკუნის მეცნიერების პირმშოა. მისი წარმოშობა, ისევე როგორც ზოგადად, დიგიტალური ჰუმანიტარიის წარმოშობა, ადამიანის ცნობიერებისა და ყოფიერების სფეროში მიმდინარე ევოლუციით არის განპირობებული. ტექნიკურმა რევოლუციამ და მისმა თანმდევმა საყოველთაო გლობალიზაციამ შეუქცევადი გახადა სამყაროს „გაციფრების“ პროცესი - დღეს წარმოუდგენელია ადამიანის კულტურულ-მენტალური და ინტელექტუალური რეალიზაცია ციფრული რესურსების გარეშე. დიგიტალური ქართველოლოგიის განვითარება თავად ჰუმანიტარული მეცნიერების შიგნით მიმდინარე კვლევითი პროცესების ტექნოლოგიზებით იყო განპირობებული და არსებითად საერთაშორისო სამეცნიერო პროექტების ფარგლებში განხორციელდა. დიდი მოცულობის ემპირიული რესურსების - დიდი მონაცემების (big data) სწრაფი დამუშავების აუცილებლობამ, ერთი მხრივ, და მონაცემთა სისტემური დამუშავების – კლასტერული ანალიზის ავტომატიზების შესაძლებლობამ, მეორე მხრივ, სრულიად ახალი პერსპექტივები გაუხსნა ჰუმანიტარული მეცნიერების ისეთ დარგებს, როგორიცაა ენათმეცნიერება, ლიტერატურათმცოდნეობა, ფილოსოფია, ისტორიოგრაფია, არქეოლოგია, ხელოვნებათმცოდნეობა, მუსიკათმცოდნეობა, კულტუროლოგია და ა.შ.

წინამდებარე ნაშრომში განხილულია დიგიტალური ქართველოლოგიის განვითარების ისტორია და პერსპექტივები, დიგიტალური კვლევის ძირითადი პრინციპები და კორპუსლინგვისტური კვლევის მეთოდები. მიმოხილულია ქართველოლოგიის დიდ მონაცემთა ბაზები, მათი შექმნის ისტორია და მნიშვნელობა. ნაშრომი განკუთვნილია ჰუმანიტარული დარგების სტუდენტებისა და საკითხით დაინტერესებული მკითხველისათვის.

რედაქტორი: პროფ. თინათინ მარგალიტაძე
რეცენზენტი: პროფ. დარეჯან თვალთვაძე

© მანანა თანდაშვილი
© მარიამ ყამარაული

ISBN 978-9941-8-3155-3

გამომცემლობა „ივერიონი“

სარჩევი

თავი პირველი – დიგიტალური ჰუმანიტარია

1. დიგიტალური ჰუმანიტარია – არსის, ცნებისა და ტერმინების განსაზღვრებისათვის 7
1.1. დიგიტალური ჰუმანიტარიის არსი 9
1.2. ჰუმანიტარია *vs.* დიგიტალური ჰუმანიტარია 14
1.3. ცნებისა და ტერმინ(ებ)ის განსაზღვრებისათვის 20
2. დიგიტალური ჰუმანიტარიის განვითარების ისტორია 25
2.1. ჰუმანიტარული კომპიუტინგიდან დიგიტალურ ჰუმანიტარიამდე 26
2.2. დიგიტალური ჰუმანიტარიის ფორმირება ინტერდისციპლინურ დარგად 29
2.3. ინტერნეტქსელის როლი დიგიტალური ჰუმანიტარიის განვითარებაში 32
2.4. დიგიტალური ჰუმანიტარიის განვითარების საკვანძო მოვლენები და რესურსები 39
2.5. მონაცემთა აფეთქების საუკუნე 43
2.6. ღია რესურსები და მათი მნიშვნელობა 46

თავი მეორე – დიგიტალური ქართველოლოგია

1. დიგიტალური ქართველოლოგიის განვითარების ისტორია, ეტაპები და ტექნოლოგიები 52
1.1. ცნებისათვის „დიგიტალური ქართველოლოგია“ 52
1.2. ქართული ენის კომპიუტერიზაცია 54
1.3. ქართველოლოგიის დიგიტალიზაციის ეტაპები 59
1.4. დიგიტალური ქართველოლოგიის კონცეპტუალიზაციისათვის 68

დანართი I – ქართველურ და კავკასიურ ენათა დიგიტალიზაციის შედეგად შექმნილი რესურსები 87

თავი მესამე – კორპუსლინგვისტიკის არსი და მნიშვნელობა ქართველოლოგიის დიგიტალიზაციისათვის

1. კორპუსული მოძრაობის ისტორია საქართველოში და დიდ მონაცემთა ბაზები 101
1.2. დიგიტალური ლექსიკოგრაფია 105
2. კორპუსლინგვისტიკის არსი, ამოცანები და მიზნები 116
3. კორპუსლინგვისტიკის კვლევის ობიექტი 120
4. რა არის კორპუსი? 122
5. კორპუსის მახასიათებლები 125

5.1. კორპუსის მოცულობა ... 125
5.2. კორპუსის სახეები ... 127
5.3. კორპუსის ტიპები ... 138
5.4. კორპუსის დაბალანსება ... 141
6. კორპუსის სტრუქტურა ... 143
6.1. სტრუქტურირებული ტექსტი კორპუსში ... 143
6.2. მეტამონაცემი ... 145
6.3. კორპუსმენეჯერი ... 148
7. კორპუსის პრერეკვიზიტები ... 149
8. კორპუსლინგვისტიკის საკვანძო ცნებები ... 153
8.1. მონაცემის ახალი სახეობა – ანოტაცია ... 153
8.2. POS-ანოტაცია ... 155
8.3. ანოტაციის დონეები ... 156
8.4. ორაზროვნების მოხსნა ... 164
8.5. მონაცემთა ვიზუალიზაცია ... 169
8.6. კოლოკაცია და კოოკურენცია ... 175
8.7. ინფორმაციის ვიზუალიზაცია გერჰარდ ჰაიერის Wortschatz-ში ... 176
9. მეთოდი კორპუსლინგვისტიკაში ... 179
9.1. კორპუსზე დამყარებული კვლევა ... 180
9.2 კორპუსით განპირობებული კვლევა ... 182

თავი მეოთხე – ქართული ენის ეროვნული კორპუსი

1. ეროვნული კორპუსის განსაზღვრებისათვის ... 187
2. ქართული ენის ეროვნული კორპუსის შექმნის ისტორია და საფუძვლები ... 188
3. ქართული ენის ეროვნული კორპუსის მოცულობა ... 189
4. ქართული ენის ეროვნული კორპუსის ეტაპობრივი განვითარება ... 189
5. ქართული ენის ეროვნული კორპუსის გამოყენება ... 192
5.1. (ქვე)კორპუსების გამოყენება ... 192
5.2. კორპუსის ტექსტების გამოტანა/ჩვენება ... 193
5.3. კორპუსში ძიება ... 195
5.4 საბაზისო ძებნა ... 196
5.5. გაფართოებული ძებნა ... 199
5.6. CQL-ის გამოყენება ძებნისას ... 202
5.7. ძებნა ტრანსლიტერაციის გამოყენებით ... 202
5.8. კონკორდანსი ... 203
6. სტატისტიკა ლინგვისტიკის სამსახურში ... 204
7. ქართული ენის ეროვნული კორპუსის სახელმწიფოებრივი მნიშვნელობა ... 218
8. ქართული ენის ეროვნული კორპუსის მეცნიერული გამოყენების სფეროები 222

თავი მეხუთე – ქართული ანბანის საერთაშორისო სტანდარტიზაცია უნიკოდში

1. უნიკოდი და ქართული ანბანის საერთაშორისო კოდირება 223
1.1. კოდირების სპეციალური პრობლემები ქართული ენის ეროვნულ კორპუსში 232
1.2. ნორმალიზაციის თანამედროვე მიდგომები 237
1.3. ვერიფიკაციის საკითხი 238
2. ხელნაწერების დიგიტალიზაცია 239
3. ვირტუალური ბიბლიოთეკის კონცეფცია 244

დანართი II – კავკასიოლოგია მაინის ფრანკფურტის უნივერსიტეტში 247

თავი პირველი
დიგიტალური ჰუმანიტარია

1. დიგიტალური ჰუმანიტარია – არსის, ცნებისა და ტერმინ(ებ)ის განსაზღვრებისათვის

დიგიტალური ჰუმანიტარია 21-ე საუკუნის მეცნიერების პირმშოა, რომლის წარმოშობაც ადამიანის ცნობიერებისა და ყოფიერების სფეროში მიმდინარე ევოლუციით არის განპირობებული და დიგიტალური ეპოქის ლოგიკურ და გარდაუვალ შედეგს წარმოადგენს. ტექნიკურმა რევოლუციამ და მისმა თანმდევმა საყოველთაო გლობალიზაციამ შეუქცევადი გახადა სამყაროს „გაციფრების" პროცესი – ადამიანის ყოველდღიურობა თანამედროვე ეპოქაში გაჯერებულია ბინარული ციფრების – ნულისა (0) და ერთის (1) კომბინაციის უსასრულო ჯაჭვით. დიგიტალურმა ეპოქამ მოიცვა არა მხოლოდ ადამიანის ყოფიერების გამოყენებითი სფერო, არამედ სრულად დაიპყრო მისი ცნობიერება – დღეს წარმოუდგენელია ადამიანის კულტურულ-მენტალური და ინტელექტუალური რეალიზება ციფრული რესურსების გარეშე.

დიგიტალური ჰუმანიტარიის წარმოშობის აუცილებლობა თავად ჰუმანიტარული მეცნიერების შიგნით მიმდინარე კვლევითი პროცესების ტექნოლოგიზებით იყო განპირობებული, რამაც **კვლევის რაოდენობრივი (კვანტიტატიური)** და **თვისებრივი (კვალიტატიური) ასპექტები** ერთ სიბრტყეში მოაქცია – დიდი მოცულობის ემპირიული რესურსების – **დიდი მონაცემების (big data)** სწრაფი დამუშავების აუცილებლობამ, ერთი მხრივ, და მონაცემთა სისტემური დამუშავების – **კლასტერული ანალიზის** ავტომატიზების შესაძლებლობამ, მეორე მხრივ, სრულიად ახალი პერსპექტივები გაუხსნა ჰუმანიტარული მეცნიერების ისეთ დარგებს, როგორიცაა ენათმეცნიერება, ლიტერატურათმცოდნეობა, ფილოსოფია, ისტორიოგრაფია, არქეოლოგია, ხელოვნებათმცოდნეობა, მუსიკათმცოდნეობა, კულტუროლოგია და ა.შ.

ემპირიული რესურსების რაოდენობრივი და თვისებრივი დამუშავების სინქრონიზაცია აუცილებლობით გულისხმობს **ინდუქციური** და **დედუქციური** კვლევის ერთ სიბრტყეში მოთავსებას. რესურსების სტატისტიკური მეთოდებით დამუშავების შედეგად მიღებული შედეგების **კლასტერულ ანალიზს**, ანუ ერთგვაროვან და განსხვავებულ ობიექტთა დაჯგუფებას აგრომელაციური და იერარქიული სტრუქტურების გენერირების მიზნით, ავტომატურად გავყავართ **თეორიულ პოსტულირებაზე.**

შესაბამისად, იქმნება სამეცნიერო კვლევის თვისებრივად ახალი პროცესი, რომელიც თავის თავში ერთდროულად მოიცავს ორ კვლევით კომპონენტს: **რაოდენობრივი სიგნიფიკატურობის დადგენას** და **სამეცნიერო პოსტულატების ავტომატურ გენერირებას**. პირველი კომპონენტი – **რაოდენობრივი სიგნიფიკატურობა** – ემპირიული რესურსების სისტემურ დამუშავებას გულისხმობს და ორი პარამეტრით განისაზღვრება:

1. **რესურსების მოცულობა** – აღნიშნული პარამეტრი სტატისტიკურად გაზომვადია და დამოკიდებულია მონაცემთა მოპოვების, შენახვისა და დამუშავების ტექნოლოგიური საშუალებების განვითარებაზე;
2. **რესურსების დამუშავების სისწრაფე** – აღნიშნული პარამეტრი იზომება **ეფექტიანობის (efficiency) ხარისხით** და დამოკიდებულია მონაცემთა დამუშავების ტექნოლოგიურ პროგრესზე.

კვლევითი პროცესის მეორე კომპონენტი – **სამეცნიერო პოსტულატების ავტომატური გენერირება** – პირველი კომპონენტის წარმატებით განხორციელებაზე არის დამოკიდებული, თუმცა დამატებით მოითხოვს მიღებული შედეგების ვერიფიცირებას. მართალია, **კლასტერიზაცია** მონაცემთა ქაოსური სიმრავლიდან ერთგვაროვან და განსხვავებულ ობიექტთა დაჯგუფებას ითვალისწინებს, რის შედეგადაც შესაძლებელი ხდება აგრომელაციური და იერარქიული სტრუქტურების გენერირება, მაგრამ ეს პროცესი თავისთავად არ გულისხმობს **მონაცემთა ავტომატურ სისტემატიზაციას**, ანუ **ქაოსური სიმრავლიდან სტრუქტურირებული სიმრავლის შექმნას**. შესაბამისად, კელეის დიგიტალური მეთოდი გულისხმობს **რაოდენობრივი** და **თვისებრივი** მეთოდების სიმბიოზს – **ფენომენების აღწერას** (*Description*) და **ფენომენების სამეცნიერო ახსნას** (*Explanation*), რაც განაპირობებს სტატისტიკურად გამყარებულ დებულებათა გენერირებას და სისტემურად განპირობებულ ახალ თეორიებს წარმოშობს. ამ გზით მიღებული თეორიები, რომლებიც დიდ მონაცემთა დამუშავების შედეგად იქმნება, ემპირიულად დადასტურებული თეორიებია და ემპირიული ადეკვატურობის მაღალი ხარისხით ხასიათდება.

თანამედროვე კვლევებისათვის აუცილებელი ეს ორი კომპონენტი ლოგიკურ ურთიერთმიმართებაში იმყოფება: **რაც უფრო მაღალია რესურსების მოცულობა (რაც თავისთავად გულისხმობს მონაცემთა მოპოვების, შენახვისა და დამუშავების ტექნოლოგიური საშუალებების მაღალ ხარისხს), მით უფრო ზუსტია ავტომატურად გენერირებადი სამეცნიერო პოსტულატები, რაც, შესაბამისად, რესურსების აღწერის ადეკვატურობის მაღალ ხარისხს უზრუნველყოფს.**

რესურსების მოპოვებისა და შენახვის ახალი ტექნოლოგიების შექმნამ, ერთი მხრივ, და რესურსების ავტომატური დამუშავების ტექნოლოგიების განვითარებამ, მეორე მხრივ, სრულიად ახალი პერსპექტივა შეუქმნა ჰუმანიტარული დარგის მკვლევრებს: მონაცემთა დიდი ბაზების (ე.წ. „დიდი მონაცემები“ *Big Data*) არსებობამ და რესურსების დიგიტალური მეთოდებით დამუშავებამ დააჩქარა და დახვეწა ჰუმანიტარულ მეცნიერებებში კვლევის პროცესი, რამაც, თავის მხრივ, ახალი სამეცნიერო დისციპლინის – **დიგიტალური ჰუმანიტარიის** (Digital Humanities) შექმნა გამოიწვია.

დღეს ბევრს დაობენ დიგიტალური ჰუმანიტარიის არსის შესახებ: მივიჩნიოთ თუ არა იგი დამოუკიდებელ სამეცნიერო დისციპლინად თუ განვიხილოთ იგი ინტერდისციპლინურ დარგად, რომელიც ჰუმანიტარული მეცნიერებისა და ინფორმატიკის ზღვარზე აღმოცენდა და ჰუმანიტარულ მეცნიერებაში დიგიტალური კვლევის თანამედროვე მეთოდების გამოყენებით შემოიფარგლება. ზოგიერთი მეცნიერი **დიგიტალურ ჰუმანიტარიას** ტრადიციული **ჰუმანიტარიის** თანამედროვე, ეპოქის მოთხოვნების გათვალისწინებით შექმნილ სინონიმად მიიჩნევს და ფიქრობს, რომ „შესაძლებელია, ათი ან ოცი წლის შემდეგ განსაზღვრება „დიგიტალური“ სედმეტად გეეჭვეენოს ჰუმანიტარულ მეცნიერებებთან მიმართებაში... და შესამჩნევიც აღარ იყოს ის ფაქტი, რომ ადამიანის შემოქმედების შესასწავლად ციფრული მეთოდები გამოიყენება“(თ.ჩ.)[1].

რა არის **დიგიტალური ჰუმანიტარია**: ტრადიციული ჰუმანიტარიის თანამედროვე სახე, რომელიც კვლევის თანამედროვე ტექნოლოგიებს იყენებს, ინტერდისციპლინური სამეცნიერო დარგი, როგორც მეცნიერების განვითარების ახალი ფორმა თუ ახალი სამეცნიერო დისციპლინა, რომელიც ტექნიკური რევოლუციის ეპოქაში სამეცნიერო მოთხოვნებისდა შესაბამისად ჩამოყალიბდა? ამ კითხვაზე პასუხის გასაცემად საჭიროა განვსაზღვროთ დიგიტალური ჰუმანიტარიის არსი, კვლევის ობიექტი და მეთოდები.

1.1. დიგიტალური ჰუმანიტარიის არსი

სამყარო, როგორც ობიექტურად მოცემული რეალობა, მეცნიერული კვლევის ძირითად საგანს წარმოადგენს და მრავალწახნაგოვანი და კომპლექსური ფენომენია. ამიტომ მეცნიერების ერთი დარგის ფარგლებში ხშირად შეუძლებელია ამა თუ იმ მოვლენის მეცნიერულად არგუმენტირებული, და იმავე დროს, მრავალწახნაგოვანი და კომპლექსური აღწერა – საჭიროა ერთდროულად ორი (ან რამდენიმე) სამეცნიერო

[1] Susan Schreibman, John Unsworth, Ray Siemens, eds., *A New Companion to Digital Humanities*. Oxford: Blackwell, 2016, XVII; რამაზ ხალვაში, დიგიტალური ჰუმანიტარიის შესავალი, ბათუმი, 2018, 49-50.

დარგის კომპეტენციის მოხმობა, რათა ბოლომდე ჩაეწვდეთ საკვლევი ობიექტის (მეტა)ფიზიკურ ბუნებასა და მისი ფუნქციონირების მექანიზმს.

მე-20 საუკუნის მეორე ნახევარი და განსაკუთრებით 21-ე საუკუნის დასაწყისი მკვეთრად გამოირჩევა მეცნიერებაში ინტერდისციპლინური დარგების მოზღვაებით. ისეთი ინტერდისციპლინური დარგები, როგორიცაა ბიოფიზიკა, რომელიც ცოცხალი სისტემების ფუნქციონირების ფიზიკურ მექანიზმს შეისწავლის, ან ბიოქიმია, რომელიც ცოცხალ ორგანიზმში მიმდინარე ქიმიურ პროცესებს სწავლობს, მეცნიერების გუშინდელი დღეა. დღეს თანამედროვე მეცნიერება გაცილებით უფრო კომპლექსურად შეისწავლის სამყაროს და იკვლევს ფაქტებისა და მოვლენების არა მხოლოდ **არსებით მხარეს ან მახასიათებელ ნიშნებს**, არამედ მათი **ფუნქციონირებისა და გარესამყაროზე ურთიერთგავლენის კანონზომიერებებს**. უახლესი ტექნოლოგიების განვითარებამ, ერთი მხრივ, და ტექნოლოგიზაციის პროცესში თეორიული კვლევების სულ უფრო მეტად გამოყენების აუცილებლობამ, მეორე მხრივ, არა მხოლოდ **ინტერდისციპლინური** (ორი სამეცნიერო დარგის მიჯნაზე აღმოცენებული), არამედ **მულტიდისციპლინური** (რამდენიმე სამეცნიერო დარგის მიჯნაზე აღმოცენებული), სოგჯერ კი **ტრანსდისციპლინური** კვლევების განვითარება მოიტანა თან.

ცნება "**ტრანსდისციპლინური**" პირველად 1970 წელს შემოიტანა შვეიცარიელმა ფილოსოფოსმა და ეპისტემიოლოგმა **ჟან პიაჟემ** (Jean Piaget). მასვე ეკუთვნის ცნების პირველი დეფინიცია. მეცნიერი მიიჩნევდა, რომ მეცნიერების განვითარება კვლევის დისციპლინური და ინტერდისციპლინური ეტაპიდან უნდა გადავიდეს კვლევის უფრო მაღალ, ტრანსდისციპლინურ ეტაპზე, რომელიც პროცესებს გლობალური სისტემის შიგნით განიხილავს დისციპლინებს შორის მკვეთრი საზღვრების გაველების გარეშე[2].

თავდაპირველად **ტრანსდისციპლინურობა** წარმოადგენდა კვლევის სტრატეგიას. იგი უგულვებელყოფდა კვლევის პროცესში დისციპლინებს შორის საზღვრების არსებობას და გულისხმობდა საკვლევი ობიექტის მულტიდისციპლინურ კვლევას მისი **ჰოლისტიკური** (მთლიანობის ამსახველი) ბუნების დასადგენად.

აღსანიშნავია, რომ გერმანულში (და შესაბამისად გერმანულენოვან ქვეყნებში) ტრანსდისციპლინურობის განსხვავებული გაგება არსებობს. **ტრანსდისციპლინურობა** გულისხმობს კვლევის სხვადასხვა ფორმების **ინტეგრაციას** და სპეციფიკური მეთოდოლოგიის გამოყენებას პრობლემის გადასაჭრელად საჭირო ცოდნის კონცენ-

[2] იხ. 1st World Congress of Trandisciplinarity, Preamble. Convento da Arrábida, Portugal, 1994, November 2-6.

ტრაციის მიზნით. „დარგი, უპირველეს ყოვლისა, აერთიანებს კვლევის პოლიტიკურ და ეკონომიკურ, სოციოლოგიურ და სოციო-ფსიქოლოგიურ, საგანმანათლებლო, კულტურულ და ენობრივ პერსპექტივებს, (...) რომელსაც განავრცობს ლიტერატურათმცოდნეობიდან შემოტანილი ჰერმენევტიკის, ესთეტიკისა და ისტორიულ რეალიებზე ორიენტირებული მედიამეცნიერების პერსპექტივა" (თ.ჩ.) – ასე ახასიათებდა ტრანსდისციპლინურობას ჰუმანიტარიაში ჰაინც პიურერი (Heinz Pürer)[3].

შესაბამისად, **ტრანსდისციპლინურობა** გერმანულენოვან სამყაროში ასოცირდება ინტეგრაციულ კვლევებთან, რომელიც დღეისათვის მეცნიერების ერთ-ერთ თანამედროვე მიმართულებას წარმოადგენს. გერმანულ ენაში იგი გადმოიცემა შესაბამისი ტერმინით Integrationswissenschaft,[4] რომელიც ქართულად შეგვიძლია ვთარგმნოთ როგორც **ინტეგრაციის შემსწავლელი მეცნიერება.**

ინტეგრაციული, ანუ ტრანსდისციპლინურ მეცნიერებას წარმოადგენს, მაგალითად, პოლიტოლოგია, რომელიც დღეს როგორც დარგი, წარმოუდგენელია ლინგვისტიკის (პოლიტლინგვისტიკა), ეკონომიკის (პოლიტიკური ეკონომიკა), ფილოსოფიის (პოლიტიდეოლოგია), იურისპრუდენციის (კანონისა და პოლიტიკის ურთიერთმიმართება ტერმინოლოგიის ასპექტში), ფსიქოლოგიის (ფსიქოლინგვისტიკა) გარეშე. ასევე, ტრანსდისციპლინურ მეცნიერებად უნდა განვიხილოთ სოციოლოგია, რომელიც საზოგადოების თანაცხოვრების ფუნქციონირების კანონზომიერებებსა და ფორმებს იკვლევს. ტრანსდისციპლინური მეცნიერების თავისებურება იმაში მდგომარეობს, რომ აქ არც მხოლოდ ცალკეული დარგისათვის დამახასიათებელი კვლევის მეთოდები და პრინციპები არის გამოყენებული და არც რომელიმე დარგი არის დომინანტური. სხვადასხვა დარგის წარმომადგენლები თანაბარუფლებიან გარემოში კოლეგებთან ერთად ქმნიან **ინტეგრაციული კვლევის მეთოდებს**, რათა მიაღწიონ **საკვლევი ობიექტის ჰოლისტიკური ბუნების გამოვლენას**. თანამედროვე მეცნიერება ტრანსდისციპლინური დარგების წარმოშობით ან უკვე არსებული ინტერდისციპლინური დარგების ტრანსდისციპლინურ დარგებად ტრანსფორმირების გზით ვითარდება. თვალსაჩინოებისთვის შეგვიძლია დავასახელოთ ახლადწარმოშობილი ინტეგრაციული დარგის რამდენიმე მაგალითი:

- **კომუნიკაციის მეცნიერება** (Communication studies / Communication sciences), რომელიც იყენებს სოციოლოგიის, ფსიქოლოგიის, ლინგვისტიკის (პრაგმატიკა),

[3] Heinz Pürer, *Publizistik- und Kommunikationswissenschaft. Ein Handbuch.* Konstanz: UVK 2003, 50.
[4] იხ. Wilfried Loth, Wolfgang Wessels, *Auf dem Weg zur Integrationswissenschaft.* In: *Grundlagen für Europa.* Vol. 7, Springer 2001, S. 7–15.

სემიოტიკის, იურისპრუდენციის, პოლიტოლოგიის და სხვა დარგების თეორიებსა და მეთოდებს;

- მედიამეცნიერება (Media studies), როგორც ტრანსდისციპლინური დარგი, იყენებს ენათმეცნიერების, ინფორმაციის თეორიის, ჟურნალისტიკის, კულტუროლოგიის, რიტორიკის (დიგიტალური რიტორიკის), ფსიქოლოგიის, პოლიტოლოგიის, საზოგადოებათმცოდნეობის, ეკონომიკის, ანთროპოლოგიის, ხელოვნების, ესთეტიკის და სხვა დარგების თეორიებსა და მეთოდებს;
- სპორტის მეცნიერება (Sports science) დღეს იყენებს ისეთი (ინტერ)დისციპლინური დარგების თეორიებსა და მეთოდებს, როგორებიცაა ბიომექანიკა, კინესიოლოგია (მოძრაობის შემსწავლელი მეცნიერება), ნევროლოგია, სპორტის დიდაქტიკა და პედაგოგიკა, სპორტის ეთიკა, სპორტული ჟურნალისტიკა, სპორტის ისტორია, სპორტის ფსიქოლოგია, ვარჯიშის მეცნიერება და სხვა. სპორტის მეცნიერების, როგორც ინტეგრაციული დარგის მიზანია სპორტის წარმომადგენელთა როგორც ფიზიკური (ბიომექანიკური), ისე ფსიქიკური, სოციალური, ეთიკური, ეკონომიკური და საზოგადოებრივი ფუნქციონირების მექანიზმის შესწავლა.

2003 წელს გოტინგენის უნივერსიტეტში საგანგებო კონფერენცია მიეძღვნა ტრანსდისციპლინურობის თემას, სადაც ცალსახად განისაზღვრა მულტიდისციპლინური, ტრანსდისციპლინური და ინტერდისციპლინური კვლევების არსი და შემუშავდა განსხვავების მადიფერენცირებელი პრინციპები ამ სამ ფენომენს შორის[5].

სამეცნიერო დარგების დისციპლინებად დაყოფის და სისტემატიზაციის საკითხს არაერთი ნაშრომი მიეძღვნა ევროპულ მეცნიერებაში, რომელმაც ერთგვარი ტერმინოლოგიური აფეთქება გამოიწვია ევროპულ სამეცნიერო სივრცეში[6]. ტრანსდისციპლინურობის გვერდით გაჩნდა ახალი ტერმინები: **დედისციპლინური, პლურიდისციპლინური, პოსტდისციპლინური, სუპრადისციპლინური**; თუმცა იმის გამო, რომ მათი მნიშვნელობა მეცნიერულად სათანადოდ არ იყო განსაზღვრული, და აქედან

---

[5] კონფერენციის მასალები იხ. კრებულში: Frank Brand, Franz Schaller und Harald Völker (Hg.), *Transdisziplinarität. Bestandsaufnahme und Perspektiven*, Beiträge zur THESIS-Arbeitstagung im Oktober 2003 in Göttingen, Universitätsverlag Göttingen, 2004.

[6] იხ. Philipp Balsiger, *Transdisziplinarität: Systematisch-vergleichende Untersuchung disziplienübergreifender Wissenschaftspraxis*. München 2005; Andrea Maihofer, *Inter-, Trans-und Postdisziplinarität. Ein Plädoyer wider die Ernüchterung*. In: Heike Kahlert, Barbara Thiessen, Ines Weller (Hg.): Quer denken – Strukturen verändern. Gender Studies zwischen den Disziplinen. Wiesbaden 2005; Gertrude Hirsch Hadorn, Holger Hoffmann-Riem, Susette Biber-Klemm, Dominique Joye, Christian Pohl, Urs Wiesmannn, Elisabeth Zemp, (Hg.): *Handbook of Transdisciplinary Research*. Dordrecht 2008.

გამომდინარე, განსხვავებული შინაარსით გამოიყენებოდა, ვერ დამკვიდრდა სამეცნიერო მიმოქცევაში.

ტრანსდისციპლინური დარგების თავისებურება, ცალკეული სამეცნიერო დარგებისაგან განსხვავებით, ძირითადად მდგომარეობს სამეცნიერო დისციპლინის შიდა სტრუქტურირების გაგებაში: სამეცნიერო დარგებისათვის გადამწყვეტია **კვლევის პროცესი** საკვლევი ობიექტის ბუნების გამოსაცნობად, მაშინ როდესაც ტრანსდისციპლინური კვლევის შემთხვევაში არსებითია **კვლევის პროცესის შედეგების ტრანსდისციპლინური ინტეგრაციის პერსპექტივიზაცია.**

კვლევის პროცესი, ტრანსდისციპლინური კვლევების შემთხვევაში, შესაბამისად, სამი საფეხურისაგან შედგება:

1. პრობლემის განსაზღვრა და სტრუქტურირება;
2. პრობლემის დამუშავება;
3. პრობლემის ტრანსდისციპლინური ინტეგრაცია.

ამგვარი მიდგომა, ანუ ტრანსდისციპლინური კვლევების განხორციელება, შედარებით ახალი ეტაპია მეცნიერებაში და ჯერჯერობით ძირითადად სამეცნიერო პროექტების ფარგლებში ხორციელდება. ტრანსდისციპლინური კვლევების მთავარი დამახასიათებელი ნიშნებია:

1. მეცნიერებასა და პრაქტიკას შორის მკვეთრი საზღვრების წაშლა;
2. რამდენიმე დარგის სპეციალისტების მჭიდრო თანამშრომლობა საგანგებო, ე.წ. **მოდერირებული დიალოგის** ფორმატში;
3. კვლევის მონიტორინგი – **კვლევის შედეგიანობის ხარისხის** (Efficiency) დადგენა.

ინტერდისციპლინური და ტრანსდისციპლინური კვლევები მეცნიერული ცოდნის პრაქტიკულ გამოყენებაზეა ორიენტირებული და, შესაბამისად, მეცნიერების განვითარებაში უმნიშვნელოვანეს ეტაპს წარმოადგენს. ტრანსდისციპლინური კვლევის ერთ-ერთ წარმატებულ მაგალითად შეიძლება დავასახელოთ ციურიხის ტექნიკური უნივერსიტეტის ETH-ის (Eidgenössische Technische Hochschule Zürich) კვლევითი ლაბორატორია **Collegium Helveticum**[7]. თუ რამდენად მნიშვნელოვანია დღეს ტრანსდისციპ-

[7] ციურიხის უნივერსიტეტის კვლევითი ლაბორატორია **Collegium Helveticum** დაარსდა 1997 წელს და მიზნად ისახავდა, ერთი მხრივ, საბუნებისმეტყველო და ტექნიკური დარგების, და მეორე მხრივ, ჰუმანიტარული და სოციალური მეცნიერების წარმომადგენელთა დიალოგს. ამაყე ლაბორატორიის ფარგლებში განხორციელდა ტრანსდისციპლინური კვლევები მეცნიერებისა და გამოყენებითი ხელოვნების დაახლოების მიზნით. ციურიხის ტექნიკური უნივერსიტეტი იმითაც არის ცნობილი, რომ უნივერსიტეტის შენობა აგებულია ცნობილი გერმანელი არქიტექტორის, დრეზდენის ოპერის შენობის ავტორის, გოტფრიდ სემპერის (Gottfried Semper) მიერ.

ლინური კვლევების განყოფილების საუნივერსიტეტო სივრცეში, ამაზე ისიც მიუთითებს, რომ აღნიშნული უნივერსიტეტი მსოფლიოში ერთ-ერთ რენომირებულ უნივერსიტეტად არის მიჩნეული, რომელმაც უნივერსიტეტების მსოფლიო საერთაშორისო კატალოგში (Shanghai-Ranking) 2013 წელს მსოფლიოში მე-20 ადგილი, ევროპაში მე-3 ადგილი, ხოლო კონტინენტურ ევროპაში პირველი ადგილი დაიკავა.

თუ გადავხედავთ დიგიტალური ჰუმანიტარიის განვითარების ისტორიას, დავრწმუნდებით, რომ როგორც კვლევის სამეტაპობრივი პროცესის, ისე კვლევის მახასიათებელი ნიშნების თვალსაზრისით, დიგიტალური ჰუმანიტარია თავისუფლად შეგვიძლია მივაკუთვნოთ ტრანსდისციპლინური სამეცნიერო დარგების ჯგუფს. ამის დასტურად გამოდგება ის მრავალრიცხოვანი ინოვაციური, პრაქტიკულ გამოყენებაზე ორიენტირებული რესურსები და ინსტიტუციები (მაგ., ენის კორპუსები, დიგიტალური ლექსიკონები, კომპიუტერული და მობილური აპლიკაციები, დიგიტალური ბიბლიოთეკები, დიგიტალური არქივები და ა.შ.), რომლებიც სხვადასხვა დარგის მეცნიერთა მჭიდრო თანამშრომლობის შედეგად შეიქმნა და საფუძველი ჩაუყარა საზოგადოების განვითარებისათვის უმნიშვნელოვანეს პროცესს: 1) **რესურსების გასაჯაროებას** და 2) **საყოველთაო წვდომას და პრაქტიკულ გამოყენებას**. სწორედ ეს მიდგომა დაედო პრინციპად **ღია რესურსების ინიციატივას**, რომელმაც ჰუმანიტარიის სწრაფი ევოლუცია გამოიწვია და განაპირობა 21-ე საუკუნის ჰუმანიტარიის – **დიგიტალური ჰუმანიტარიის** შექმნა.

1.2. ჰუმანიტარია *vs.* დიგიტალური ჰუმანიტარია

დიგიტალური რესურსების, როგორც კვლევის თანამედროვე აუცილებელი კომპონენტის, შექმნა და დიგიტალური დამუშავების მეთოდოლოგიის განვითარება, არსებითად განასხვავებს დიგიტალურ ჰუმანიტარიას ტრადიციული ჰუმანიტარიისაგან. თუ ერთმანეთს შევადარებთ **ტრადიციული ჰუმანიტარიისა** და **დიგიტალური ჰუმანიტარიის** არსს, თვალსაჩინო გახდება განსხვავება მათ შორის. ეს განსხვავება არსებითად კვლევის ობიექტის, როგორც **„პირველადი“** და **„მეორადი“ რესურსის** გაგებას ეფუძნება.

ჩვენი ცოდნა კვლევის ობიექტის შესახებ ყოველთვის არ ემყარება თავად ობიექტის ჩვენეულ „აღქმას“ კოგნიტურ დონეზე (ანუ აღწერას და ანალიზს). იგი ხშირად გაშუალებულია იმ პოსტულატებით, რომელიც სამეცნიერო კვლევებს ეფუძნება. შესაბამისად, ჩვენ უმეტესად „ვიცით“ ის, რაც თეორიულ პოსტულატებში არის მოცემული, რომელიც თავის მხრივ კონკრეტულ მეცნიერთა (ან სამეცნიერო სკოლის წარმომადგენელთა) თეორიული ჩარჩოს ფარგლებში „დასტურდება“. მისი

ვერიფიცირების საშუალება, ხშირად შეუძლებელია ობიექტური და სუბიექტური ფაქტორების გამო. ენის მაგალითზე რომ განვიხილოთ: ენის გრამატიკა ენას სისტემურად აღწერს, რამდენადაც იკელევს კანონზომიერებებს, რომელიც ემპირიულად დასტურდება და განაზოგადებს მას წესების მეშვეობით. ენის ლექსიკონი არის ენაში მოცემული ლექსიკური მარაგის რეპრეზენტაცია, რომელიც მეტნაკლებად ასახავს ენის ლექსიკურ ფონდს მისი ფუნქციური სემანტიკის გათვალისწინებით (განმარტებითი ლექსიკონის შემთხვევაში) ან წარმოგვიდგენს თითოეული ლექსიკური (ან ფუნქციური) ერთეულის (ლექსიკურ ან ფუნქციურ) ეკვივალენტს მეორე ენაში (ორენოვანი ლექსიკონის შემთხვევაში). შესაბამისად, ჩვენი ცოდნა ემყარება არა პირველად (ენის ემპირიულ მასალას), არამედ მეორად რესურსს (სტრუქტურირებულ ცოდნას). დიგიტალური ჰუმანიტარიის განვითარება, უპირველეს ყოლისა, თანამედროვე ტექნოლოგიებით გამართულ ემპირიულ მონაცემთა ბაზების შექმნასთან ასოცირდება, რომელიც იძლევა რაოდენობრივი მნიშვნელობის დადგენისა და სამეცნიერო პოსტულატების ავტომატური გენერირების საშუალებას, რაც კვლევის რაოდენობრივ და თვისებრივ მეთოდებს ერთ სიბრტყეში მოაქცევს და კვლევის ეფექტიანობას საგრძნობლად ზრდის. კვლევის, როგორც პროცესის სწორედ ეს არსებითი მხარე განაპირობებს განსხვავებას ტრადიციულ ჰუმანიტარიასა და დიგიტალურ ჰუმანიტარიას შორის. შედარებისათვის განვიხილოთ ეს ორი ოდენობა თვისებრივი მახასიათებლების მიხედვით.

ტრადიციულ ჰუმანიტარიას ახასიათებს შემდეგი ნიშნები:

ა) პირველად რესურსებზე წვდომის (დროული და სივრცული) სირთულე;

ჰუმანიტარული დარგის კვლევის ობიექტები, მაგალითად, ხელნაწერი ან ბეჭდური რესურსები, არქეოლოგიური არტეფაქტი, არქიტექტურული ნაგებობა, ხელოვნების ნიმუში და ა.შ., ისტორიულად არის თავმოყრილი ერთ ადგილას, გეოგრაფიულად მთელ მსოფლიოშია განფენილი და მათზე წვდომა მეცნიერული კვლევების ჩასატარებლად სივრცული და დროული სირთულეების გადალახვას მოითხოვს. საქართველოს მაგალითზე რომ განვიხილოთ, ქართული ხელნაწერების დიდი უმრავლესობა, რომელიც ილია ჭავჭავაძის თაოსნობითა და საქართველოში წერა-კითხვის გამავრცელებელი საზოგადოების ძალისხმევით შეგროვდა, ძირითადად თავმოყრილია კ. კეკელიძის სახელობის საქართველოს ხელნაწერთა ინსტიტუტში (დღეს კ. კეკელიძის სახელობის საქართველის ხელნაწერთა ეროვნული ცენტრი), როგორც ქართული კულტურული მემკვიდრეობის წერილობითი განძის უდიდესი ნაწილი; თუმცა ქარ-

თულ ხელნაწერთა გარკვეული ნაწილი უცხოეთში, სხვადასხვა ქვეყნის ბიბლიოთეკებში, მუზეუმებში, საცავებში, სამონასტრო და კულტურულ ცენტრებში ინახება. შესაბამისად, ამგვარ რესურსებზე წვდომა და მათი დამუშავება გარკვეულ სირთულეს წარმოადგენს დროული და სიერცული დაშორების გამო, რასაც ემატება ადგილობრიე ინსტიტუციებთან ურთიერთობის სიძნელე[8]. იგიე შეიძლება ითქვას ქართული არქიტექტურული ძეგლების შესახებაც, რომელთა ნაწილი დღეს საქართველოს ფარგლებს გარეთ მდებარეობს, მაგალითად, თურქეთში, და მათზე მუშაობა (ასომეა, ფოტოგრაფირება, რესტავრაცია/რეკონსტრუქცია) გარკვეულ სირთულეებთან არის დაკავშირებული.

ბ) პირველადი რესურსების ხელით დამუშავება;

ტრადიციული ჰუმანიტარიის ფარგლებში რესურსების დამუშაეება ხელით ხორციელდებოდა. ენათმეცნიერების მაგალითზე რომ განვიხილოთ, სალექსიკონო მასალის მოპოვების ტრადიციული მეთოდი – დაბარათება – მკელევრისაგან მოითხოვდა ხანგრძლივ და განმეორებად შრომას. მაგ., ტექსტის სიმფონია-ლექსიკონის შექმნის შემთხვევაში ერთი წინადადების ამობარათება ხდებოდა იმდენჯერ, რამდენი სიტყვაც იყო წინადადებაში. ასევე, არქიტექტურული ძეგლების აზომეის შემთხვევაში მონაცემთა დამუშაევება ხორციელდებოდა ხელით, რაც დიდ დროს მოითხოვდა და ტექნიკური და ადამიანური შესაძლებლობის ფონზე არ გამორიცხავდა ცდომილებას.

გ) პირველად რესურსებზე ერთჯერადი წვდომა;

სხვადასხვა გეოგრაფიულ გარემოში, როგორც საქართველოს, ისე უცხოეთის საცავებში, ბიბლიოთეკებსა და არქივებში (ან სხვა მსგავს ინსტიტუციებში) დაცული რესურსების კვლევა ძირითადად საგანგებო ექსპედიციების ფარგლებში ხორციელდებოდა, რესურსზე ერთჯერადი წვდომის გზით. სამეცნიერო ჰიპოთეზის დაზუსტების ან ახალი დებულების შესამოწმებლად საჭირო ხდებოდა დროსა და სიერცეში ხელახალი გადაადგილება და რესურსის ხელახალი დამუშავება.

---

[8] აქ ვგულისხმობთ ხელნაწერებზე მუშაობის ნებართვას, რომელიც ხშირად ძნელი მისაღებია, განსაკუთრებით სამონასტრო კომპლექსებში დაცული ხელნაწერების შემთხვევაში. თუმცა უნდა აღინიშნოს, რომ ხელნაწერებზე თავისუფალი წვდომა ზოგჯერ საქართველოში, კერძო საკუთრებაში არსებული ხელნაწერების შემთხვევაშიც არ არის ადვილი, განსაკუთრებით სვანეთში დაცული ხელნაწერების შემთხვევაში. მაგალითად, სოფელ ლახამულაში დაცული ხელნაწერი, რომელიც საკმაოდ არის დაზიანებული და რესტავრაციას მოითხოვს, მიუწვდომელი აღმოჩნდა ჩვენი სამეცნიერო ჯგუფისათვის, რომელიც მიზნად ისახავდა ხელნაწერის დიგიტალიზაციას, ვინაიდან ადგილობრივი მაცხოვრებლები მას საკრალურ ნივთად მიიჩნევენ და არა წერილობით ძეგლად.

დ) სამეცნიერო პოსტულატების გადამოწმების (ვერიფიცირების) სირთულე;

მომხმარებელი ძირითადად სამეცნიერო ლიტერატურის (ანუ მეორადი რესურსის) მეშვეობით ეცნობოდა ამა თუ იმ წყაროს, რაც გამორიცხავდა სამეცნიერო პოსტულატების გადამოწმების (ვერიფიცირების) შესაძლებლობას რესურსის შიგნით. ხშირ შემთხვევაში პირველად რესურსებზე წვდომა შეუძლებელი იყო.

ე) ემპირიული მასალის დაზუსტების შემთხვევაში ხელახალი გამოცემის აუცილებლობა;

ნებისმიერი ბეჭდური გამოცემა (ტექსტი, ალბომი) შეიძლება შეიცავდეს ისეთ უზუსტობებს, როგორიცაა მაგალითად, კორექტურული შეცდომები, ბეჭდვითი პროცესის შედეგად მოდიფიცირებული ფერები, დაუზუსტებელი ზომები, რომლის გასწორებაც ბეჭდური გამოცემის შემთხვევაში მხოლოდ ხელახალი გამოცემით არის შესაძლებელი, რაც დამატებით ადამიანურ და ფინანსურ რესურსებს მოითხოვს და ხშირად შეუძლებელია ობიექტური ან სუბიექტური მიზეზების გამო, მაშინ, როდესაც დიგიტალური გამოცემების შემთხვევაში შეცდომების ჩასწორება, კორექტირება/რედაქტირება ან განერცობა/გადამუშავება სირთულეს არ წარმოადგენს.

ვ) სამეცნიერო სივრცეში მიმოქცევის შეზღუდულობა;

რესურსების სამეცნიერო სივრცეში მიმოქცევის შეზღუდულობა სამი ფაქტორით იყო განპირობებული: **იდეოლოგიით** (მაგალითად, საბჭოთა პერიოდში აკრძალული „ბურჟუაზიული" სამეცნიერო ლიტერატურა და ზოგჯერ მთელი დარგებიც კი), **აკადემიური მონოპოლიზაციით** (კვლევის თემები ხშირად პერსონიფიცირებული იყო და ასოცირდებოდა კონკრეტულ ადამიანებთან, რაც ზღუდავდა მისი, როგორც თავისუფალი სამეცნიერო კვლევითი რესურსის მიმოქცევას) და **რესურსზე წვდომის შეუძლებლობით**, რაც განპირობებული იყო როგორც სუბიექტური, ისე ობიექტური ფაქტორებით (მაგ., ჩაკეტილი პოლიტიკური სივრცე, საომარი მოქმედებები, ქვეყნებს შორის დიპლომატიური ურთიერთობების არარსებობა და ა.შ.). **დიგიტალური ჰუმანიტარიის** ფარგლებში კი თანამედროვე რესურსები, იმდენად, რამდენადაც დიგიტალური ფორმით არსებობს და ამასთან, ადვილად მოპოვებადია ღია **რესურსების კონცეფციიდან** გამომდინარე, კვლევის თავისუფალ რესურსებს წარმოადგენს, მარტივად ხელმისაწვდომი და მრავალჯერადი გამოყენებისაა.

ტრადიციული ჰუმანიტარიისაგან განსხვავებით **დიგიტალური ჰუმანიტარია** ხასიათდება შემდეგი ნიშნებით:

ა) პირველად რესურსებზე წვდომის სიმარტივე და პირველადი რესურსების მოპოვების სისწრაფე;

დიგიტალური ბაზების არსებობა ამარტივებს პირველადი რესურსების მოპოვებას და ეს სიმარტივე გამოიხატება არა მარტო რესურსების მოპოვებისათვის საჭირო დროში (დიგიტალური ბაზების გამოყენების შემთხვევაში თქვენ შეგიძლიათ რამდენიმე წამში მოიპოვოთ თქვენთვის საჭირო რესურსი), არამედ სივრცული ბარიერის გადალახვაში – რესურსების მოსაპოვებლად აღარ არის საჭირო მათი რეალური ადგილსამყოფლის მონახულება, დიგიტალური რესურსი თქვენ შეგიძლიათ ვირტუალურად „გამოიძახოთ“ თქვენს კომპიუტერში და იკვლიოთ იგი თქვენი ინტერესის შესაბამისად. ამგვარად, თქვენ იყენებთ ე.წ. „ვირტუალურ ბიბლიოთეკებსა და საცავებს/არქივებს“. შესაბამისად, თქვენ გიწევთ მუშაობა არა გეოგრაფიულ სივრცეში, არამედ ვირტუალურ გარემოში.

ბ) პირველადი რესურსების მანქანური (ელექტრონული) დამუშავება;

რესურსების ხელით დამუშავება დიგიტალური ჰუმანიტარიის ფარგლებში შეცვლილია კომპიუტერული დამუშავებით, რაც გიზოგავთ დროსა და ენერგიას და მიღებული შედეგებიც გაცილებით ზუსტი და სისტემურია, ვიდრე ხელით დამუშავების დროს. მაგალითად, არქიტექტურული ძეგლების დოკუმენტირება-რეკონსტრუირება (ფოტოგრაფირება/ასომვა/რეკონსტრუქცია) დიგიტალურად ხორციელდება საგანგებო აპარატურისა და პროგრამული უსრუნველყოფის გამოყენების გზით, რაც უსრუნველყოფს მიღებული შედეგების სიზუსტესა და ადეკვატურობას.

გ) პირველად რესურსებზე მრავალჯერადი წვდომა;

დიგიტალური ჰუმანიტარიის რესურსი რეალურად არსებული რესურსის ვირტუალური ანალოგია, რომელიც ელექტრონულ მატარებლებზე არის შენახული და მრავალჯერადი გამოყენებისაა. მასზე წვდომა შეგიძლიათ განახორციელოთ იმდენჯერ, რამდენჯერაც მოითხოვს თქვენი საკვლევი თემა ან სამეცნიერო ინტერესები.

დ) სამეცნიერო პოსტულატების გადამოწმების (ვერიფიცირების) შესაძლებლობა;

დიგიტალური რესურსები ძირითადად აღჭურვილია ორიგინალური ფოტო-, ვიდეო-, აუდიო- და სხვა ვიზუალური მასალით, რაც საშუალებას გაძლევთ გადაამოწმოთ ელექტრონული რესურსების სიზუსტე ან გამოიყენოთ რესურსები ინტერდისციპლინური კვლევებისათვის. მაგალითად, წარწერების კორპუსის რესურსებს თან ერთვის შესაბამისი ფოტომასალა, რაც საშუალებას იძლევა აღნიშნული რესურსი

ხელახლა დაამუშაოთ სიზუსტის დასადგენად ან გამოიყენოთ ინტერდისციპლინური კვლევებისათვის, მაგალითად, „გადაამოწმოთ“ რესურსი არა მარტო ლინგვისტური, არამედ პალეოგრაფიული თვალსაზრისითაც.

ე) ემპირიული ბაზის (შედარებითი) სისრულე;

დიგიტალურ ჰუმანიტარიაში კვლევის შედეგების სიზუსტე და სისტემურობა განპირობებულია ემპირიული ბაზის შედარებითი სისრულით, ვინაიდან კვლევა დიგიტალურ ჰუმანიტარიაში ეყრდნობა **დიდ მონაცემთა ბაზას**. ამასთან, გასათვალისწინებელია, რომ დიდი მონაცემის შემთხვევაშიც საქმე შეიძლება გვქონდეს **შედარებით სისრულესთან**, განსაკუთრებით ეს ეხება ღია კორპუსებს. სტატიკური კორპუსების შემთხვევაში ემპირიული მასალის სისრულის მიღწევა შესაძლებელი და აუცილებელიცაა.

ვ) კვლევის სისტემურობა;

დიდ მონაცემთა ბაზის დამუშავება, ერთი მხრივ, და კვლევის თანამედროვე ტექნოლოგიების გამოყენება, მეორე მხრივ, თავისთავად უზრუნველყოფს სისტემურობას, ვინაიდან კვლევის რესურსების დამუშავება მანქანურად ხორციელდება წინასწარ შემუშავებული კომპლექსური წესების გამოყენებით.

ზ) გამართული მეთოდურობა;

დიგიტალურ ჰუმანიტარიაში კვლევა არა მხოლოდ სისტემურად ხორციელდება, არამედ გამართული მეთოდურობითაც ხასიათდება. კორპუსული კვლევების დროს ძირითადად გამოიყენება ორი უმნიშვნელოვანესი მეთოდი: **კორპუსზე ორიენტირებული კვლევის მეთოდი** და **კორპუსით განპირობებული კვლევის მეთოდი**. ეს ორივე მეთოდი ერთმანეთს განავრცობს და უზრუნველყოფს კვლევის შედეგად მიღებული დასკვნების ადეკვატურობას რესურსთან მიმართებაში.

თ) რედაქტირების სიმარტივე და ელექტრონული განახლებადობა

აღმოჩენილი უზუსტობების გასწორება ან მასალის შემდგომი დამუშავება-განვრცობა დიგიტალური ჰუმანიტარიის ფარგლებში მარტივად ხორციელდება. ელექტრონული რესურსები განახლების (update) სიმარტივით ხასიათდება და ნაკლებ პრობლემას წარმოადგენს ეკონომიკური თვალსაზრისითაც – განახლებული ვერსიები ბეჭდურად აღარ უნდა გამოიცეს; ეს რესურსები ვირტუალურ ბიბლიოთეკებშივე „სწორდება“.

ი) მობილურობა (სამეცნიერო სივრცეში მიმოქცევის თავისუფლება);

მკვლევართა მიერ მოპოვებულ რესურსებზე სამეცნიერო კუთვნილების საკითხის გადაწყვეტამ და კვლევითი რესურსების „გასაჯაროებამ" 21-ე საუკუნის ჰუმანიტარია არსებითად შეცვალა – დღეს მსოფლიო სამეცნიერო საზოგადოება **„ღია რესურსების"** პრინციპის გამოყენებით ეითარდება. ნებისმიერ ადამიანს შეუძლია მოიპოვოს საჭირო რესურსი კონკრეტული სივრციდან გაუსვლელად.

ზემოთქმულიდან გამომდინარე, დიგიტალური ჰუმანიტარის განსაზღვრება ამგვარად უნდა ჩამოვაყალიბოთ: დიგიტალური ჰუმანიტარია ჰუმანიტარული მეცნიერებებისა და ინფორმატიკის მიჯნაზე აღმოცენებული ტრანსდისციპლინური სამეცნიერო დარგია, რომელიც კვლევის დროს სისტემატურად იყენებს დიგიტალურ რესურსებსა და კომპიუტერულ ტექნოლოგიებს ჰუმანიტარულ დარგებსა და კულტუროლოგიაში თეორიული კვლევების საწარმოებლად. დიგიტალური ჰუმანიტარია ეითარდება „ღია რესურსების" შექმნისა და გამოყენების გზით. დიგიტალური რესურსების თანამედროვე ტექნოლოგიებით დამუშავების შედეგად მიღებული თეორიული დასკვნები, კვლევის მეთოდებიდან გამომდინარე, სიზუსტით, კვლევის ადეკვატურობითა და სისტემურობით გამოირჩევა. დიგიტალური ჰუმანიტარია – ჰუმანიტარიისა და ინფორმაციული ტექნოლოგიებისაგან შექმნილი სინერგეტიკული დისციპლინაა და 21-ე საუკუნის მეცნიერების განვითარების გარდაუვალ შედეგს წარმოადგენს.

13. ცნებისა და ტერმინ(ებ)ის განსაზღვრებისათვის

სამეცნიერო დისციპლინა, რომელზეც ჩვენ ვსაუბრობთ, ახალი დისციპლინაა და მისი აღმნიშვნელი ტერმინი – დიგიტალური ჰუმანიტარია – პირდაპირი თარგმანია ინგლისური ტერმინისა **digital humanities**. ამ ტერმინის სინონიმად ქართულში აგრეთვე გამოიყენება ტერმინი ციფრული ჰუმანიტარია. ამ ორი ტერმინის ურთიერთმიმართების შესახებ მოგვიანებით საგანგებოდ ვიმსჯელებთ. ამჯერად ვისაუბრებთ ცნებისა და ტერმინის ურთიერთმიმართების შესახებ.

სამეცნიერო ტერმინოლოგიური ლექსიკონების მიხედვით ცნება არის აზროვნების ერთ-ერთი ლოგიკური ფორმა, განსოგადების უმაღლესი დონე. მისი მეშვეობით ხდება რაიმე საგნის, მოვლენის, ობიექტური სინამდვილის და ა.შ. შემეცნება. შესაბამისად, ცნება შეიძლება განვიხილოთ, როგორც შემეცნების ოპერატორი. ცნების შემუშავება ხორციელდება განსაზღვრების საშუალებით, რომლის სინონიმადაც ქართულენოვან სამეცნიერო ლიტერატურაში ასევე გვხვდება ლათინური ტერმინი დეფინიცია. დეფინიციის მეშვეობით ჩვენ გამოვთქვამთ ჩვენს აზრს ან დამოკიდებულებას

აღსანიშნი ობიექტის – საგნის, მოელენის, ობიექტური სინამდვილის შესახებ ორი არსებითი ნიშნის მიხედვით: ეს არის ცნების მოცულობა და შინაარსი. ცნების მახასიათებლების ერთობლიობა ქმნის ცნების შინაარსს, ხოლო შინაარსით განსაზღვრული საგნების რაოდენობა – ცნების მოცულობას. არსებობს კანონზომიერი ურთიერთდამოკიდებულება ამ ორ მახასიათებელს შორის: რაც უფრო ფართოა ცნება მოცულობით, მით უფრო ღარიბია ის შინაარსით და, პირიქით – შინაარსით მდიდარი ცნება უფრო მეტად ისწრაფვის კონკრეტულობისაკენ და, მაშასადამე, ნაკლებია მისი მოცულობაც[9].

ტერმინი არის სიტყვა ან სიტყვათა შეთანხმება, რომელსაც იყენებენ რომელიმე სპეციალური დარგის ცნების ან სპეციალური საგნის აღსანიშნავად. ეს არის ცნების ვერბალური განსხეულება, რომელიც ზუსტად გადმოსცემს მეცნიერების, ტექნიკის, ხელოვნების ან რომელიმე სხვა დარგის გარკვეული ცნების შინაარსსა და მოცულობას.

თუ ცნების განსაზღვრებისას მოგვეთხოვება ვიყოთ თანმიმდევრულები, **ტერმინის** შერჩევის დროს მეტი თავისუფლება გვაქვს მოცემული. ტერმინის შემუშავება კონვენციის (შეთანხმების) შედეგად ხორციელდება. ტერმინი შეიძლება შევარჩიოთ ენაში უკვე არსებული ლექსიკიდან, ან ვისესხოთ სხვა ენიდან. მთავარია, ცნების განსაზღვრებისა და კონვენციურად მიღებული მისი აღმნიშვნელი ტერმინის დადგენის შემდეგ აღარ შევცვალოთ **ცნების ოპერატორი** და მიღებული ტერმინი მსჯელობის (ან უბრალოდ, კომუნიკაციის) პროცესში მუდმივად ერთი და იმავე მნიშვნელობით გამოვიყენოთ. კიდევ ერთხელ უნდა დავაზუსტოთ, რომ **ტერმინი** არის მეცნიერული ოპერატორი, ანუ მეტაენისათვის სპეციალურად შემოტანილი სიტყვა, რომელიც ყოველთვის კონვენციურად მიღებული წესით უნდა გამოიყენებოდეს მოცემული დარგის ან თეორიის ფარგლებში.

ინგლისურენოვან სამყაროში **დიგიტალური ჰუმანიტარიის** აღსანიშნად ასევე გამოიყენება ტერმინი **e-Humanities**, რომელიც მომიჯნავე ტერმინის, კერძოდ **e-Sciences**-ის ანალოგიით შეიქმნა. ორივე ტერმინი – **Digital humanities** და **e-Humanities** – შედარებით ახლად დანერგილი ტერმინებია და ჩაენაცვლა მანამდე არსებულ და ამჟამად მოძველებულ ტერმინებს: **Computing in the humanities** და **Humanities computing**.

თუ აღნიშნული ტერმინების სიერცულ და შინაარსობრივ კომპონენტებს ჩავუღრმავდებით, ვნახავთ, რომ დიგიტალური ჰუმანიტარიის თავდაპირველ ტერმინოლოგიურ გააზრებებში აქცენტი დასმული იყო არა ცოდნის **ჰუმანიტარულ შინაარს-**

[9] იხ. Civil ენციკლოპედიური ლექსიკონი http://www.nplg.gov.ge/gwdict/index.php?a=index&d=5

ზე. არამედ კომპიუტერული ტექნოლოგიების ფუნქციონირებაზე ჰუმანიტარულ დარგებში, რის გამოც ტერმინი „**Computing in the Humanities**“ სულ მალე შეიცვალა ტერმინით „**Humanities Computing**“, თუმცა ცნების შინაარსი ამით დიდად არ შეცვლილა. უბრალოდ „ჰუმანიტარია“ და „კომპიუტინგი“ შესიტყვებაში „ჰუმანიტარული კომპიუტინგი“ შედარებით გათანაბრდა, თუმცა „ჰუმანიტარია“ ტერმინის სტრუქტურაში მაინც მსაზღვრელის ფუნქციით იქნა გამოყენებული და ლოგიკურ-შინაარსობრივი მახეილი კვლავ „კომპიუტინგზე“ დარჩა. აღნიშნული ტერმინების შემოტანამ და სამეცნიერო მეტაენაში მათმა გამოყენებამ მნიშვნელოვანი როლი შეასრულა „დარგობრივი ინფორმატიკების“ (ბიოინფორმატიკა, ქიმიოინფორმატიკა, საინჟინრო ინფორმატიკა, ბიზნესინფორმატიკა, სოციოინფორმატიკა, ხელოვნური ინტელექტი, კომპიუტერული ლინგვისტიკა, კორპუსლინგვისტიკა, ეკოინფორმატიკა, ნეიროინფორმატიკა) განვითარებასა და სხვადასხვა ჰუმანიტარულ მეცნიერებაში კომპიუტერული ტექნოლოგიების ფართო გამოყენებაში.

დარგის განვითარებასთან ერთად, დროთა განმავლობაში, ცხადი გახდა, რომ ტერმინი **Humanities Computing** ვეღარ ასახავდა არტეფაქტებისა და ტექსტების გაციფრების როგორც მასშტაბს, ისე შესაძლებლობებს. ცნების ხელახალი განსაზღვრების შედეგად 2004 წელს შეიქმნა ახალი ტერმინი „დიგიტალური ჰუმანიტარია“, რომელშიც, წინა ტერმინებისაგან განსხვავებით, აქცენტი კომპიუტინგიდან ჰუმანიტარიაზე იქნა გადატანილი.

ახალი ტერმინის მიხედვით **დიგიტალური ჰუმანიტარია** არის ჰუმანიტარული დარგების ერთობლიობა, რომელიც სისტემატურად ქმნის და იყენებს დიგიტალურ რესურსებსა და კვლევის დიგიტალურ მეთოდებს თეორიული კვლევების განსახორციელებლად ჰუმანიტარულ დარგებსა და კულტუროლოგიაში. ახალი ტერმინის შემოღებით არა მარტო გაფართოვდა ცნების შინაარსი, არამედ მოცულობაც. ცნების თანამედროვე დეფინიციის მიხედვით **Humanities Computing**-ი, რომელიც ინსტრუმენტების, ინფრასტრუქტურების, სტანდარტებისა და კოლექციების შექმნას გულისხმობს, გაერთიანდა **ჰუმანიტარულ მეცნიერებათა ბლოგინგთან** (მედიაქსელური სტრუქტურების წარმოება და დიგიტალური თანამშრომლობა) და **მულტიმოდალურ ჰუმანიტარიასთან** (მეცნიერული ინსტრუმენტების, მონაცემთა ბაზებისა და თანამშრომლობის ქსელური სტრუქტურების გაერთიანება, ასევე ვიდეო და აუდიო მედიასაშუალებების გამოყენება)[10].

[10] Tara McPherson. *Introduction: Media Studies and the Digital Humanities* // Cinema Journal. 2009, 119-123.

როგორც ზემოთ უკვე აღვნიშნეთ, ქართულ ენაში დღეისათვის ორი ტერმინი გამოიყენება: **დიგიტალური ჰუმანიტარია** და **ციფრული ჰუმანიტარია**. ხშირად ამ ტერმინებს სინონიმურ ტერმინებად განიხილავენ, თუმცა უპირატესობას მაინც ამ უკანასკნელს ანიჭებენ იმ მოტივით, რომ „დიგიტალური" უცხო სიტყვაა ქართულში, „ციფრული" კი – ქართულ ენაში უკვე დამკვიდრებული. დღევანდელი მონაცემებით ინტერნეტში ამ ტერმინების გამოყენების სიხშირე „დიგიტალური ჰუმანიტარიის" სასარგებლოდ მეტყველებს: დიგიტალური ჰუმანიტარია – 709, ციფრული ჰუმანიტარია – 472[11]. თუ ჩვენ ამ პრობლემას ქართველოლოგიასთან მიმართებაში განვიხილავთ, ცალსახად უნდა ვთქვათ, რომ საქმე გვაქვს ორ განსხვავებულ ოდენობასთან, რომელიც ერთმანეთისაგან განსხვავდება როგორც შინაარსის, ისე მოცულობის მიხედვით.

დიგიტალური ჰუმანიტარიის ეროვნულ მეცნიერებაში შემოტანას წინ უსწრებდა ქართველოლოგიის გაციფრების პროცესი. თანამედროვე ქართულ სამეცნიერო სივრცეში ტერმინი **დიგიტალიზაცია** ხშირად **რესურსების გაციფრებას** გულისხმობდა – იქმნებოდა არსებული ლექსიკონების ელექტრონული ვარიანტები, მიმდინარეობდა ელექტრონული ტექსტური ბაზების შექმნა, ბიბლიოთეკების კატალოგების გაციფრება; მეტიც, ბოლო წლებში ბეჭდური რესურსების გაციფრების „ბუმი" შეუქცევად პროცესად იქცა. მაგალითისათვის შეგვიძლია დავასახელოთ საქართველოში ჯერ კიდევ გასული საუკუნეში დაწყებული პროექტი **გაზეთების ელექტრონული კორპუსი** (ეს რესურსი ცნობილია როგორც **open text**), ქართველი ავტორებისა და მათი ლიტერატურული ნაწარმოებების ელექტრონული ბაზა **lib.ge**, საქართველოს პარლამენტის ეროვნული ბიბლიოთეკის მიერ უკანასკნელ წლებში შექმნილი ციფრული ბიბლიოთეკა, რომელიც დღეისათვის ტიპობრივად და შინაარსობრივად განსხვავებულ რესურსებს მოიცავს: ა) ციფრული ბიბლიოთეკა „იუერიელი", რომელიც გვთავაზობს საქართველოს პარლამენტის ეროვნულ ბიბლიოთეკაში დაცული წიგნების, ჟურნალების, გაზეთების და ა.შ. ელექტრონულ ასლებს, ბ) მ. მამარდაშვილის სახელობის ციფრული კოლექცია, რომელიც მოიცავს ქართული კლასიკური ლიტერატურის ყველა ცნობილი წარმომადგენლის ნაწარმოებების ტექსტებს. აქვეა გამოცემების სრული ტექსტები სამოქალაქო განათლებაზე, ეკონომიკაზე, სამართალზე და სხვ., გ) ელექტრონული ჟურნალები და სრულტექსტოვან გამოცემათა ბაზები ეროვნული ბიბლიოთეკის ლოკალურ ქსელში და დ) Moodstanza – ლექსები – მობილური აპლიკაცია ანდროიდის მომხმარებლებისთვის, რომლის საშუალებითაც შეგიძლიათ წაი-

[11] სტატისტიკური მონაცემის თარიღი: 02.01.2021

კიოხოთ XIX-XX საუკუნის ქართველი კლასიკოსების პოეზიის ნიმუშები[12].

მიუხედავად იმისა, რომ აღნიშნული რესურსები ხელმისაწვდომია ფართო საზოგადოებისათვის, იგი ჯერ კიდევ არ წარმოადგენს **დიგიტალური ჰუმანიტარიის რესურსებს**, ვინაიდან მათ არ გააჩნიათ უმნიშვნელოვანესი მახასიათებლები: **ტექსტები არ არის სტრუქტურირებული, არ გააჩნიათ შიდასისტემური საძიებო სისტემა, არ არიან აღჭურვილი მეტამონაცემებით (ფილტრის პერსპექტივით), რაც შეუძლებელს ხდის მათ გამოყენებას თანამედროვე კვლევებისათვის, განსაკუთრებით ინტერ-** ან ტრანსდისციპლინური კვლევების შემთხვევაში. რესურსების გაციფრების მოცემულ ეტაპზე ნიშანდობლივი იყო ის ფაქტორი, რომ მასშტაბურად განხორციელდა რესურსების ელექტრონული „გადაკოპირება“ ისე, რომ არ მოხდა მათი თანამედროვე თეორიული და მეთოდოლოგიური თვალსაზრისით გადამუშავება. დღეისათვის არსებული გაციფრებული ქართველოლოგიური რესურსების უმრავლესობა არ არის ერთმანეთთან დაკავშირებული კომპლექსური კვლევების განსახორციელებლად. აქედან გამომდინარე, ზემოთ განხილული ტერმინები **ციფრული ქართველოლოგია** და **დიგიტალური ქართველოლოგია** ჩვენ არ მიგვაჩნია სინონიმებად: ტერმინი **ციფრული ქართველოლოგია** მიემართება ქართველოლოგიის ელექტრონულ რესურსებს, ტერმინი **დიგიტალური ქართველოლოგია** კი მიემართება ციფრული რესურსების დიდ მონაცემთა საგანგებოდ სტრუქტურირებულ ბაზებს, აღჭურვილს თანამედროვე ტექნოლოგიური ინსტრუმენტებით, რომლებიც ქართველოლოგიაში დიგიტალური კვლევის საშუალებას იძლევიან.

[12] იხ. https://www.nplg.gov.ge/geo/Digital_Library

2. დიგიტალური ჰუმანიტარიის განვითარების ისტორია

დიგიტალური ჰუმანიტარიის განვითარების ისტორია, როგორც ზემოთ უკვე აღვნიშნეთ, არაერთხაზოვანია და დარგის ცნების (და შესაბამისად ამოცანების) არსის გაგებას უკავშირდება. დარგის ჩამოყალიბება განაპირობა, ერთი მხრივ, ჰუმანიტარულ დარგებში კვლევის პროცესის თვისებრივმა ცვლილებამ, რაც კვლევის რესურსების სიდიდესა და რესურსების გადამუშავების სისწრაფის აუცილებლობით იყო განპირობებული[13], მეორე მხრივ კი კომპიუტერული ტექნოლოგიების სწრაფმა განვითარებამ. კომპიუტერული ტექნოლოგიების შესაძლებლობა – დროის ძალიან მცირე მონაკვეთში განახორციელოს დიდი მოცულობის ინფორმაციის დამუშავება, სწრაფად აიტაცა კაცობრიობამ და დასაბამი მისცა კომპიუტერული ტექნოლოგიების პრაქტიკულ გამოყენებას როგორც ყოველდღიურ, ისე სამეცნიერო სფეროში.

კომპიუტერის პროტოტიპის, პირველი გამომთველელი მანქანის შექმნის იდეა ეკუთვნის ინგლისელ მათემატიკოსსა და გამომგონებელს **ჩარლზ ბებიჯს** (Charles Babbage). ფართო საზოგადოებისათვის იგი ცნობილია როგორც მათემატიკოსი და გამომგონებელი. მან პირველმა შეიმუშავა მართვადი მანქანის – **კომპიუტერის** კონცეფცია, რომელსაც ანალიტიკური ოპერაციების შესრულება შეეძლო, რის გამოც იგი ცნობილია სახელით „ანალიტიკური მანქანა" (Analytical Engine)[14]; თუმცა ბებიჯის ინტერესები და მოღვაწეობის სფერო ბევრად სცილდებოდა ერთი დარგის ფარგლებს – იგი იყო ფილოსოფოსი და პოლიტეკონომისტი. ბებიჯს ძალიან კარგად ესმოდა მის მიერ შექმნილი მექანიზმის მნიშვნელობა კაცობრიობის განვითარების ისტორიაში. თავის წიგნში „ფრაგმენტები ფილოსოფოსის ცხოვრებიდან" (*Passages from the Life of a Philosopher*) იგი წერდა: „როგორც კი შეიქმნება ანალიტიკური მანქანა, ის აუცილებლად გახდება მეცნიერების სამომავლო კურსის განმსაზღვრელი. მისი დახმარებით მიღებული შედეგები დასვამს კითხვას: რა გაანგარიშების შედეგად შეიძლება მივიღოთ ეს შედეგები მანქანის მეშვეობით უმოკლეს დროში?" (თ.ჩ.)[15]. მიუხედავად იმისა, რომ ბებიჯი ვერ მოესწრო თავისი იდეის პრაქტიკულ განხორციელებას,

---

[13] აქ შეგვიძლია, უპირველეს ყოვლისა, დავასახელოთ იტალიელი იეზუიტი მღვდლის **რობერტო ბუზას** მიერ დასახული ამოცანა – თომა აქვინელის შემოქმედებისა და მის მიმდევართა ტექსტების ბაზის შექმნა.

[14] ანალიტიკური მანქანის პირველი აღწერა ბებიჯმა 1834 წელს გამოაქვეყნა, თუმცა ფინანსური პრობლემების გამო მანქანის მხოლოდ რამდენიმე ნაწილის კონსტრუირება შეძლო.

[15] *"As soon as an Analytical Engine exists, it will necessarily guide the future course of the science. Whenever any result is sought by its aid, the question will then arise – By what course of calculation can these results be arrived at by the machine in the shortest time?".*

მას მაინც საპატიო ადგილი უჭირავს დიგიტალური ჰუმანიტარიის განვითარებაში: იგი გამოთვლითი ტექნიკის მამის სახელით შევიდა ისტორიაში.

2.1. ჰუმანიტარული კომპიუტინგიდან დიგიტალურ ჰუმანიტარიამდე

გამოთვლითი ტექნიკის განვითარების ახალი ეტაპი XIX საუკუნეში დაიწყო და ამერიკაში დემოგრაფიული პრობლემის, კერძოდ მოსახლეობის აღწერის გადაჭრას ემსახურებოდა. ამერიკელი ინჟინრისა და გამომგონებლის **ჰერმან ჰოლერითის** (Herman Hollerith) მიერ შექმნილმა ელექტროგამომთვლელმა მანქანამ უზრუნველყო ამერიკაში 1888 წელს ჩატარებული მოსახლეობის აღწერის სისწრაფე. ჰოლერითის გამოგონების წარმატებით გამოყენების შემდეგ გამომგონებელმა თავისი გამოგონების კომერციალიზაცია გადაწყვიტა და დააარსა კომპანია **Tabulating Machine Co**, რომლის საფუძველზეც 1924 წელს შეიქმნა კორპორაცია **International Business Machines (IBM)**, რომელიც დღემდე განაგრძობს არსებობას და რომელმაც უზრუნველყო არნახული ტექნოლოგიური ნახტომი მექანიკური გამომთველელი მანქანიდან ციფრულ ტექნოლოგიამდე[16].

ტექნოლოგიური რევოლუციის შემდგომი ეტაპი უკავშირდება ბრიტანელი მათემატიკოსისა და გამომგონებლის **ალან ტიურინგის** (Alan Turing) სახელს. ისევე, როგორც ბებიჯი, ტიურინგიც ფართო ინტერესებისა და ცოდნის მატარებელი იყო. უპირველეს ყოვლისა, იგი იყო ლოგიკოსი, ინფორმატიკოსი და კრიპტოანალიტიკოსი. სწორედ მის მიერ იქნა შემუშავებული დიდი ბრიტანეთის სახელმწიფო კრიპტოანალიტიკურ ცენტრში ერთ-ერთი პირველი ელექტროგამომთველელი მანქანა მსოფლიოში (**Colossus computer**), რომელსაც შეეძლო წამში 25 ათასი ნიშნის გაანალიზება და რომლის მეშვეობითაც 1944 წელს მოხერხდა ფაშისტური გერმანიის რადიომაუწყებლობით გადაცემულ ინფორმაციებში კოდირებული შეტყობინებების გაშიფრა. ტიურინგის მიერ მომდევნო წლებში შექმნილი კომპიუტერის **Automatic Computing Engine (ACE)** დიზაინი თანამედროვე კომპიუტერების პროტოტიპს წარმოადგენს. საგულისხმოა, რომ ალან ტიურინგის სახელს ატარებს ინფორმატიკაში ყველაზე მნიშვნელოვანი და პრესტიჟული პრემია **Turing Award**, რომელიც ყოველწლიურად გაიცემა ინფორმატიკის დარგში საგანგებო მიღწევების აღსანიშნავად და თავისი მნიშვნელობით უტოლდება ნობელის პრემიას. ალან ტიურინგის სახელს ატარებს ასევე საგანგებო ტესტი (Turing-Test), რომელიც გამოიყენება ადამიანური და ხელოვნური

[16] ამ ქვეთავთან დაკავშირებით დაწვრილებით იხ.: რ. ხალვაში, დიგიტალური ჰუმანიტარიის შესავალი, ბათუმი, 2018.

ინტელექტის შესაძლებლობების შესადარებლად – ტესტის მიზანს წარმოადგენს იმის დადგენა, თუ რამდენად შესაძლებელია ადამიანური აზროვნების ხელოვნური ინტელექტით შეცვლა. ალან ტიურინგის თეორიული ნაშრომები საფუძვლად დაედო თეორიულ ინფორმატიკას და მნიშვნელოვნად განაპირობა ამ დარგის განვითარების დინამიკა.

კომპიუტერების სისტემური და ფუნქციური დახვეწის პროცესი რამდენიმე ეტაპს მოიცავდა. კომპიუტერების პირველი თაობის მანქანა – პირველი თაობის კომპიუტერი ENIAC-ი (Electronic Numerical Integrator and Computer), პენსილვანიის უნივერსიტეტში შეიქმნა ჯონ მოკლისა (John Mauchly) და ჯონ ეკერტის (John Eckert) მიერ.



ელექტრონულ-სხივური მილაკებისაგან შემდგარი კომპიუტერი ფიზიკურად გიგანტურ მანქანას წარმოადგენდა: მანქანა 27 ტონაზე მეტს იწონიდა, მოცულობით 2,4 x 0,9 x 30 მეტრის იყო და 170 კვადრატული მეტრი ფართობი ეკავა. მანქანაში ჩამონტაჟებული იყო 70 000 რეზისტორი, 10 000 ელექტრული კონდენსატორი, 1 500 რელე, 7 200 კრისტალური დიოდი და 17 468 ელექტრონული ლამპა.

ENIAC-ი 1946 წლის 14 თებერვალს წარუდგინეს ავტორებმა საზოგადოებას და თითქმის 10 წელი გამოიყენებოდა როგორც სახელმწიფო, ისე სამხედრო დანიშნულებით. 1955 წლის 2 ოქტომბერს 11:45 საათზე კომპიუტერი გამორთეს. მისი რამდენიმე დეტალი გამოფენილია ამერიკის ისტორიის ეროვნულ მუზეუმში ვაშინგტონში. პირველი თაობის კომპიუტერის ხსოვნის პატივსაცემად მისი სახელი დაერქვა ასტეროიდს ნომრით 22977.

მეორე თაობის კომპიუტერების შექმნა დაკავშირებულია უნგრული წარმოშობის ამერიკელი მათემატიკოსის ჯონ ფონ ნიუმენის (John von Neumann) სახელთან. მისმა ნაშრომებმა მათემატიკის სხვადასხვა სფეროში, როგორიცაა მათემატიკური ლოგიკა, კვანტური მექანიკა, ფუნქციური ანალიზი, გადამწყვეტი როლი შეასრულა. მასვე ეკუთვნის თამაშებში ოპტიმალური სტრატეგიის შესწავლის მათემატიკური მეთოდის დამუშავება. იგი სამართლიანად ითვლება ინფორმატიკის მამად.

ნიუმენის კონცეფციის სიახლე კომპიუტერის ლოგიკურ სტრუქტურირებაში მდგომარეობდა. მისი წარმოდგენით კომპიუტერი უნდა შედგებოდეს **შესაყვანი (input)** და **გამოსაყვანი (output)** მოწყობილობებისაგან, ცენტრალური პროცესორისა (გამო-

თვლების საწარმოებლად) და მეხსიერების ბლოკისაგან. სწორედ ეს სამი კომპონენტი, რომელიც კომპიუტერის მუშაობის ლოგიკურ სქემას წარმოადგენს, უდევს საფუძვლად კომპიუტერების თანამედროვე თაობას. ნიუმენის აღნიშნულმა კონცეფციამ მძლავრი ბიძგი მისცა კომპიუტერული სისტემების განვითარებას, თუმცა კომპიუტერიზაციის პროცესს სერიოზულად აფერხებდა დანადგარების სიდიდე და სიძვირე. პრობლემა გადაიჭრა ორი უმნიშვნელოვანესი სიახლის შემუშავების შედეგად: ელექტრონული ლამპა ჩანაცვლდა ტრანზისტორით და შეიქმნა ინტეგრალური სქემა (ე.წ. მიკროსქემა). ეს უკანასკნელი ამოცანა ერთმანეთისაგან დამოუკიდებლად ერთდროულად გადაჭრეს რობერტ ნოისმა (Robert Noyce) და ჯეკ კილბიმ (Jack Kilby).

კომპიუტერების მესამე თაობის ეგზემპლარი IBM-360 1964 წელს წარმოადგინა ფირმა IBM-მა. 1969 წელს მიკროპროცესორის გამოგონებამ, ერთი მხრივ, რაც მარშინ ჰაფის (Marcian Hoff) სახელთანაა დაკავშირებული, და დუგლას ენგელბარტის (Douglas Engelbart) მიერ მაუსის შექმნამ, მეორე მხრივ, არა მარტო მინიატურობა შესძინა და დახვეწა მანქანის დიზაინი, არამედ კომპაქტური და მოსახმარებლად მოქნილი გახადა კომპიუტერი.

კომპიუტერული ინდუსტრიის განვითარების მომდევნო ეტაპის დასაწყისად შეიძლება ჩაითვალოს 1976 წელს სტივენ ჯობსის (Steven Jobs) მიერ კომპანია Apple-ის დაარსება, რომელმაც 1977 წელს პერსონალური კომპიუტერების მასობრივი წარმოება დაიწყო. 1978 წელს გამოშვებულ Apple II-ს უკვე გააჩნდა ინტეგრირებული კლავიატურა, ფერადი გრაფიკა, ხმა, 4 კილობაიტი მეხსიერება, მეხსიერების დისკი და პლასტმასის კორპუსი. 1984 წელს Apple-მა წარმოადგინა კომპიუტერი Macintosh-ი, რომლის მოხმარებასაც აადვილებდა პროგრამული უზრუნველყოფა და სხვადასხვა დანართით შევსებული ოპერაციული სისტემა. პროგრამული უზრუნველყოფის სფეროში ბილ გეიტსის (Bill Gates) გამოჩენამ სათავე დაუდო ახალ ერას – 1975 წელს მან ჰარვარდელ მეგობართან პოლ ალენთან (Paul Allen) ერთად ახალი კორპორაცია Microsoft-ი დააარსა[17]. 1981 წელს IBM-ის მიერ გამოშვებული პერსონალური კომპიუტერი PC (Personal Computer) Microsoft-ის ოპერაციულ სისტემას DOS (Disk Operating System) იყენებდა.

კომპიუტერიზაციის შემდგომ აფეთქებას ხელი შეუწყო WWW (World Wide Web) ქსელის შექმნამ, რომელიც 1991 წელს ამოქმედდა და მსოფლიო საზოგადოება კომუნიკაციის ახალი საშუალებით – ინტერნეტით უზრუნველყო. ელექტრონულმა კომუნიკაციამ ინტერაქტიული ხასიათი შეიძინა, ხოლო კიბერსივრცემ – სოციალური

[17] იხ. www.microsoft.com

ელფერი, რამაც ინტერნეტის მომხმარებელთა რიცხვის არნახული ზრდა გამოიწვია. განვითარდა ინტერნეტში ჩართვის უამრავი საშუალება: თანამგზავრული კავშირი, სატელეფონო კავშირი, ფიჭური კავშირი, სპეციალური ოპტიკურ-ბოჭკოვანი ხაზები. ინტერნეტი თანამედროვე ცხოვრების განუყოფელი ნაწილი გახდა და სრულიად მოიცვა ჩვენი ყოველდღიური ცხოვრება. თუ 2008 წელს ინტერნეტის მომხმარებელთა რაოდენობა 21% აღწევდა, 2017 წლის მონაცემებით, მომხმარებელთა რაოდენობა 49%-მდე გაიზარდა. საქართველოში ინტერნეტის მომხმარებელთა რიცხვი დღეისათვის მოსახლეობის 60,7% პროცენტს შეადგენს[18].

2.2. დიგიტალური ჰუმანიტარიის ფორმირება ინტერდისციპლინურ დარგად

ჰუმანიტარულ მეცნიერებებში და კულტუროლოგიაში კომპიუტერის შესაძლებლობების გამოყენება გასული საუკუნის 60-იან წლებში დაიწყო. ერთ-ერთი პირველი კონფერენცია, რომელიც ამ თემას მიეძღვნა, გაიმართა ავსტრიაში 1962 წელს. კონფერენციის შედეგები მოგვიანებით ცალკე წიგნადაც გამოიცა[19]. კონფერენციაზე განხილული საკითხები მოიცავდა კომპიუტერების ზოგადი გამოყენების შესაძლებლობასა და პერსპექტივებს ჰუმანიტარულ მეცნიერებებში, ერთი მხრივ, და კულტუროლოგიური მონაცემების ელექტრონულ დამუშავებას, მეორე მხრივ. კონფერენციის პრაქტიკული შედეგი ისიც იყო, რომ სულ მალე შეიქმნა ამ საკითხით დაინტერესებულ სპეციალისტთა წრე, რომლებიც განიხილავდნენ ამ და მსგავს თემებს ინტერდისციპლინურ კონტექსტში. წამყვანი თემა, რომელიც დიგიტალური ჰუმანიტარიის ფორმირების პირველსავე წლებში გაჩნდა, იყო კომპიუტერის დახმარებით ტექსტების ავტომატური ანალიზი. ამ პერიოდისათვის უკვე არსებობდა უამრავი ტექსტური კორპუსი, რომელიც იდეალური იყო ისეთი შრომატევადი პრაქტიკული სამუშაოების ავტომატურად განსახორციელებლად, როგორიცაა სიტყვების სიხშირის გამოთვლა, ან ინდექსის ავტომატური გენერირება. კომპიუტერების წარმატებით გამოყენების მეორე მნიშვნელოვანი სფერო იყო სერიული ისტორიული წყაროების ავტომატური დამუშავება, როგორიცაა მაგ., დაბადების ჩანაწერებისა ან გარდაცვალების მოწმობების სისტემატიზაცია. ამ რაოდენობრივი ტიპის ისტორიოგრაფიამ პიკს 1970-იან და 1980-იან წლებში მიაღწია, როდესაც მრავალრიცხოვანი მონაცემები, განსაკუთრებით მო-

[18]იხ. http://www.internetworldstats.com/; რ. ხალვაში, დიგიტალური ჰუმანიტარიის შესავალი, ბათუმი, 2018, გვ. 17-18.

[19] Dell H. Hymes (Hrsg.), *The Use of Computers in Anthropology*, (Studies in General Anthropology; 2), London 1965.

სახლეობის ისტორიის და სოციოლოგიის სფეროებში, სწორედ კომპიუტერის საშუალებით მუშაობდნენ. მონაცემთა კომპიუტერული დამუშავება ამ პერიოდში აღიქმებოდა არა როგორც მეთოდოლოგიური ინოვაცია, არამედ როგორც დამხმარე რელიეფური საშუალება: „კომპიუტერის გამოყენებაში თვისებრივად ახალი არაფერია, მაგრამ კომპიუტერი საშუალებას იძლევა უფრო ზუსტი და სწრაფი პასუხები გაეცეს ტრადიციულ შეკითხვებს“ (თ.ჩ.)[20].

ამ პერიოდში მიმდინარე აქტივობები განიხილებოდა საერთო ქუდის – ჰუმანიტარული კომპიუტინგის ქვეშ. როგორც ფილოლოგიაში, ასევე ისტორიულ მეცნიერებებში თითოეულს ჰქონდა საკუთარი სპეციფიკური ინფრასტრუქტურა კონფერენციებით, პროფესიონალური საზოგადოებებითა და ჟურნალებით. 1972 წელს სოციოლოგებმა და ისტორიკოსებმა შექმნეს არაფორმალური სამუშაო ჯგუფი კიოლნის უნივერსიტეტში, რათა განეხილათ სოციალურ მეცნიერებებსა და ისტორიოგრაფიაში კვლევის მეთოდების დაახლოების საკითხი. 1974 წელს ბრაუნშვაიგში შედგა რაოდენობრივი მეთოდებით დაინტერესებულ ისტორიკოსთა პირველი შეხვედრა, რომლის შედეგადაც 1975 წელს შეიქმნა სამუშაო ჯგუფი ისტორიულ და საზოგადოებრივ მეცნიერებებში რაოდენობრივი მეთოდების გამოყენების მიზნით. ამ ჯგუფის ბაზაზე კიოლნში 1977 წელს დაარსდა ისტორიული სოციოლოგიის კვლევითი ცენტრი, რომელიც შემდეგში შეუერთდა GESIS-ის ლაიბნიცის სოციალურ მეცნიერებათა ინსტიტუტს[21]. ამ პერიოდის მოქმედი საზოგადოებებიდან განსაკუთრებული აღნიშვნის ღირსია ისტორიისა და გამოთვლების ასოციაცია, რომელიც 1987 წელს დაარსდა და სისტემატურად გამოსცემდა სამეცნიერო ჟურნალს „ისტორია და გამოთვლები“.

ფილოლოგიისაგან განსხვავებით, რომელიც საუკუნოვან ტრადიციებს ითვლის, ისტორიული მეცნიერება ძირითადად ნაციონალური ისტორიოგრაფიის განვითარების გზით ვითარდებოდა და ნაკლებად იყო საერთაშორისო კვლევებით გაჯერებული. ჰუმანიტარული კომპიუტერიზაციის განვითარების ადრეულ ეტაპზე ისტორიოგრაფია ძირითადად რაოდენობრივი მეთოდების გამოყენებით იყო დაინტერესებული. ისტორიოგრაფიის ინტერნაციონალიზაცია და აქტიური ინტეგრაცია დიგიტალური ჰუმანიტარიის განვითარების პროცესში 90-იანი წლებიდან იწყება და მსოფლიო საინტერნეტო ქსელის შექმნასთან არის დაკავშირებული. აღსანიშნავია, რომ ისტორიკოსები უფრო აქტიურად ჩაებნენ საინტერნეტო რესურსების შექმნასა და

[20] Konrad H. Jarausch, *Möglichkeiten und Probleme der Quantifizierung in der Geschichtswissenschaft*, in: ders. (Hrsg.), Quantifizierung in der Geschichtswissenschaft. Probleme und Möglichkeiten, Düsseldorf 1976, S. 11-30. hier S. 13. William O. Aydelotte, *Quantifizierung in der Geschichtswissenschaft*, in: Hans-Ulrich Wehler (Hrsg.), Geschichte und Soziologie, 2. Aufl. Königstein 1984, S. 259-282.

[21] https://www.gesis.org/home

ინტეგრაციის პროცესში, ვიდრე სხვა ჰუმანიტარული დისციპლინები. თუმცა, აქვე უნდა აღინიშნოს, რომ ციფრული ტექნიკისადმი დომინანტური ინტერესი, უპირველეს ყოვლისა, ნაკლებად იყო ჰეშმარიტად კულტურულ-სამეცნიერო და ძირითადად საბიბლიოთეკო და კომუნიკაციური ხასიათის ამოცანებით შემოისაზღვრებოდა. ამ პერიოდში ყურადღება გამახვილდა ისტორიული ხასიათის ინფორმაციის მოპოვებაზე, ინფორმაციის დამუშავების შესაძლებლობებზე, რესურსთა გაცვლასა და სპეციალისტთა დარგობრივი ხასიათის კომუნიკაციაზე.

ისტორიულ მეცნიერებათა სპეციალისტების დარგობრივმა კომუნიკაციამ საკმაოდ დიდი ბიძგი მისცა და დინამიზმი შესძინა ისტორიული მეცნიერების ინტეგრაციას დიგიტალური ჰუმანიტარიის განვითარების ადრეულ პერიოდში. ამის დასტურია აშშ–ში ყოველისმომცველი საფოსტო ქსელის შექმნა, სახელწოდებით H-Net (Humanities and Social Sciences Online)[22], რომელიც წარმოადგენს ჰუმანიტარული და სოციალური მეცნიერებების სპეციალისტთა ონლაინპორტალს. იგი 1993 წელს შეიქმნა და მისი ტექნიკური ინფრასტრუქტურა განთავსებულია მიჩიგანის უნივერსიტეტში. H-Net-ის შექმნის იდეა ეკუთვნის რიჩარდ იენსენს (Richard Jensen), რომელმაც ქსელის კონცეფცია 1992 წლის დეკემბერში გააჟღერა, 1993 წლის თებერვალში კი პირველი დარგობრივი სიები წარმოადგინა, რომლის რაოდენობა წლის ბოლოსათვის უკვე 21-ს შეადგენდა. აღნიშნული დარგობრივი სიები შეიცავდა სხვადასხვა ინფორმაციებს – პუბლიკაციებს, რეცენზიებს, განაცხადებს კონფერენციებისა და სხვა აკადემიური აქტივობების შესახებ, ვაკანსიებისა და საცხოვრებელი ბინების გაქირავების შესახებაც კი.

H-Net-ის შექმნამ ხელი შეუწყო სამეცნიერო სივრცეში აკადემიური და ინფორმაციული რესურსების სწრაფ მიმოქცევას და სპეციალისტთა კომუნიკაციას. H-Net- ის პრეტენზია თავდაპირველად უფრო ფართო იყო, ვიდრე მხოლოდ კომუნიკაციის სისწრაფე: ინიციატორებს სურდათ დიგიტალურ ჰუმანიტარიაში კოლეგების კომპეტენციის გაღრმავება და ისტორიულად რელევანტური ონლაინრესურსების გამოყენების ხელშეწყობა. პირველი სია მიეძღვნა ურბანიზაციასა და ურბანიზაციის ისტორიის კვლევას (H-URBAN). შემდგომში სიების თემატიკა გაიზარდა: შეიქმნა თემატური სია, რომელიც ეძღვნებოდა, მაგალითად, ქალთა და გენდერული ისტორიის საკითხებს, ჰოლოკოსტის კვლევებს[23] და ა.შ. დღეს ქსელი ითვლის 180-მდე სხვა-

[22] http://www.h-net.org/

[23] Richard Jensen, *Internet's Republic of Letters:* H-Net for Scholars, 1997 (http://web.archive.org/web/20000903084512/http://members.aol.com/dann01/whatis.html).

დასხვა სპეციალობის ჩამონათვალს. 1996 წელს დაიწყო ქსელის გერმანულენოვანი ოფიშორული პროგრამა აკრონიმით H/Soz/Kult, რომელიც სოციოლოგიისა და კულტუროლოგიის ისტორიის საკითხებს მოიცავს[24].



ინგლისურენოვანი სიებისგან განსხვავებით, გერმანულენოვან ვერსიაში თემატური დიფერენციაცია არ განხორციელებულა. დღეს ამ სიაში დაახლოებით 20000 აბონენტია გაერთიანებული და ყოველწლიურად აქვეყნებს ათასობით რესურსს: ნაშრომებს, რეცენზიებს, ტექნიკურ მიმოხილვებს, საკონფერენციო მასალებს, ვაკანსიებს და განცხადებებს ისტორიულ მეცნიერებათა დარგში დაგეგმილი აქტივობების შესახებ.

2.3. ინტერნეტქსელის განვითარება და მისი როლი დიგიტალური ჰუმანიტარიის განვითარებაში

ისტორიულ-სამეცნიერო გილდიის მიერ მსოფლიო ქსელის გამოყენების ისტორიაში რამდენიმე პერიოდი უნდა გამოვყოთ. პირველი პერიოდი, რომელიც 90-იანი წლების შუა ხანებში დაიწყო, შეიძლება დავახასიათოთ როგორც დაგროვების პერიოდი. ამ პერიოდში წინა პლანზე წამოწეულია ინფორმაციის შეგროვება, გაცოცრება და ქსელში გამოქვეყნება. ინფორმაციის ხელმისაწვდომი წყაროების არჩევანი ამ პერიოდში საკმაოდ მოკრძალებულია და პირველ რიგში მოიცავს საბიბლიოთეკო კატალოგებს. ამ პერიოდის განმავლობაში ქსელში ნაკლებად გვხვდება

[24] http://hsozkult.geschichte.hu-berlin.de/; იხ. Rüdiger Hohls/Peter Helmberger (Hrsg.), *Humanities-Net Sozial- und Kulturgeschichte (H-Soz-u-Kult). Bilanz nach 3 Jahren* (Historical Social Research; 24/3), Köln 1999; Klaus Gantert, *H-Soz-u-Kult – Informationsdienst für die Geschichtswissenschaften*, in: Geschichte in Wissenschaft und Unterricht 62 (2011), H. 11/12, S. 645-650.

რეტროსპექტულად გაციფრებული („რეტროდიგიტალიზებული“) მასალა, რაც გარკვეულწილად ყოფილი ქსელური კავშირების დაბალი სიჩქარით შეიძლება აიხსნას[25].

საუკუნის ბოლოს შეიმჩნეოდა მსოფლიო ქსელის გამოყენების ახალი კონცეფცია: მკვლევართა და კვლევითი ინსტიტუტების მიერ საკუთარი კვლევის შედეგების **ინტერნეტსივრცეში გამოქვეყნება**. მიუხედავად იმისა, რომ ეს დაკავშირებული იყო დიდ ტექნიკურ ძალისხმევასთან ინფორმაციის მიმწოდებლის მხრიდან, ძალიან მიმზიდველი და პოპულარული გახდა, ერთი მხრივ, ახალი მკითხველების მოზიდვისა და მეორე მხრივ, პუბლიკაციის ახალი ფორმების განვითარების თვალსაზრისით. მაგალითისათვის შეგვიძლია დავასახელოთ პროექტი **pastperfect.at**[26], რომელიც შემუშავდა ისტორიკოსთა, პროგრამისტთა და გრაფიკული მხატვრების ჯგუფის მიერ ვენის უნივერსიტეტში და მოიცავდა XV-XVI საუკუნეების დოკუმენტებს. პროექტის ფარგლებში შეიქმნა დიდი მოცულობის **ჰიპერტექსტური კორპუსი**, რომელიც შეიცავდა რამდენიმე ასეულ ტექსტს ათასობით ბმულით[27].

ამავდროულად, მრავალმა მსხვილმა ინსტიტუციამ დაიწყო მათ საკუთრებაში არსებული რესურსების გაციფრება და მისი ქსელში განთავსება დაინტერესებული პროფესიონალი საზოგადოებისთვის ხელმისაწვდომობის მიზნით. ამას ხელი შეუწყო მიზნობრივი დაფინანსების პოლიტიკამ სახელმწიფოს მხრიდან. მაგალითად, გერმანიაში სამეცნიერო ფონდის **DFG**-ს (Deutsche Forschungsgemeinschaft) მიერ, აშშ-ში ჰუმანიტარულ მეცნიერებათა ეროვნული ფონდის **NEH** (National Endowment for the Humanities) მიერ დაფინანსებული პროექტების განხორციელების შედეგად შეიქმნა დიდი რაოდენობით ისტორიული წყაროების კოლექციები, როგორიცაა მაგალითად, „კომპაქტური მეხსიერება“ (Compact Memory)[28], რომელიც ებრაელთა ისტორიის ამსახველ მასალებს მოიცავს, „ღია საზოგადოების არქივი“ (OSA)[29], რომელიც ბუდაპეშტში შეიქმნა და ცენტრალური ევროპის თანამედროვე ისტორიის ამსახველ მასალებს მოიცავს, „ონლაინნაშრომები პედაგოგიკაში“ (Scripta Paedagogica Online)[30], რომელშიც შეკრებილი და გაერთიანებულია რესურსები განათლების ისტორიის შესახებ. ამ

---

[25] Ellen Collins/Michael Jubb, *How do Researchers in the Humanities Use Information Resources?*, in: Liber Quarterly 21 (2012), H. 2, S. 176-187.

[26] http://pastperfect.univie.ac.at

[27] Jakob Krameritsch, *Geschichte(n) im Netzwerk. Hypertext und dessen Potenziale für die Produktion, Repräsentation und Rezeption der historischen Erzählung*, Münster u.a. 2007.

[28] http://www.compactmemory.de

[29] http://www.osaarchivum.org

[30] http://bbf.dipf.de/digitale-bbf/scripta-paedagogica-online

პერიოდს გლობალური გამოქვეყნების პერიოდს უწოდებენ.

2000-იანი წლების მეორე ათწლეულში ქსელის მიმდინარე გამოყენების რეჟიმებს კიდევ ერთი ასპექტი დაემატა. Web2.0-ის სახით გაჩნდა თვისებრივად ახალი, თავისუფალი ინტერაქტიული ქსელის სერვისების სერია, რომელმაც არა მარტო გაზარდა მომხმარებელთა რაოდენობა, არამედ თავად მომხმარებელი აქცია ქსელური ინფორმაციის თანაშემქმნელად. ამ პროცესმა გამოიწვია სოციალური ელემენტის გააქტიურება, ერთი მხრივ, მეორე მხრივ კი ხელი შეუწყო ქსელის ინტერაქტიურობის ხარისხის ამაღლებას. Web2.0-ის განვითარებამ ქსელი აქცია სოციალურ ქსელად, მედია კი – სოციალურ მედიად. Web2.0-ის მიერ შემოთავაზებულმა პროგრამულმა და ტექნიკურმა შესაძლებლობებმა სრულიად ახალი ფუნქციით აღჭურვა ქსელი: მასალების გამოქვეყნება, მასალების გაზიარება და სოციალიზაცია[31]. შესაბამისად, **გლობალური გამოქვეყნების პერიოდი ჩაანაცვლა სოციალური ინტერაქტიულობის პერიოდმა.**

ქსელის განვითარების შემდგომ ეტაპზე რამდენიმე ქსელური პროექტი განხორციელდა, რომელიც სოციალური მედიის სხვადასხვა ელემენტების ინტეგრაციის მაღალ ხარისხს გვიჩვენებს. ამის მაგალითად შეგვიძლია დავასახელოთ პორტალი Zeitgeschichte-online, რომელიც 2004 წლიდან არსებობს და ჩამოყალიბდა, როგორც გერმანულენოვანი სამყაროს ცენტრალური პლატფორმა მედიასივრცეში, რომელსაც ერთობლივად ახორციელებს პოტსდამის თანამედროვე ისტორიის ცენტრი ZZF (Zentrum für Zeithistorische Forschung) და ბერლინის სახელმწიფო ბიბლიოთეკა – SBB (Staatsbibliothek zu Berlin), კერძოდ, პრუსიის კულტურული მემკვიდრეობის დეპარტამენტი. პლატფორმის მომსახურება მოიცავს თანამედროვე და ისტორიული ინსტიტუტების ცნობარებს, რეცენზიებს და ღონისძიებების გრაფიკს.

სოციალური მედიის ელემენტების თვალსაზრისით განსაკუთრებით აღსანიშნავია 1953 წლის ამსახველი თემატური პორტალი www.17juni53.de,[32] რომელიც არა მარტო თავად აქცევნებს 1953 წლის 17 ივნისის აჯანყების ამსახველ რესურსებს, არამედ მოუწოდებს მომხმარებლებს მონაწილეობა მიიღონ პორტალის რესურსების განვრცობაში – მიაწოდონ ამ მოვლენასთან დაკავშირებული საკუთარი მასალა.

---

[31] საკითხის დაწვრილებით გასაცნობად იხ. რესურსები: Peter Haber/Jan Hodel, *Geschichtswissenschaft und Web 2.0.* Eine Dokumentation (= The hist.net Working Paper Series, Basel, 2010) http://www.histnet.ch/repository/hnwps/hnwps-02.pdf; Michael Nentwich,René König. *Cyberscience 2.0. Research in the Age of Digital Social Networks* (Interaktiva; 11), Frankfurt a.M. 2012.

[32] http://www.17juni53.de



აღნიშნული თემატური პორტალისაგან განსხვავებით „www.chronik-der-mauer.de“ წინა პლანზე აყენებს მულტიმედიურ ელემენტებს[33]. ამ პორტალზე შეიძლება ატვირთოთ ბერლინის კედლის დაცემის არა მარტო ფოტომასალა ან პირადი მოგონებები, არამედ ვიდეომასალები, ხმოვანი ჩანაწერები და ა.შ.



თემატური პორტალებიდან უნდა დავასახელოთ აგრეთვე პლატფორმა www.ungarn1956.de, რომელიც საგრძნობლად განსხვავდება ბერლინის კედლის პორტალისაგან, ვინაიდან იგი უფრო მეტად სპეციალისტებისათვის განკუთვნილი პორ-

[33] http://www.chronik-der-mauer.de

ტალია, რომელიც მომხმარებელს 1956 წელს უნგრეთის აჯანყების თემაზე არსებულ ამომწურავ ლიტერატურასა და ონლაინმასალებს სთავაზობს.



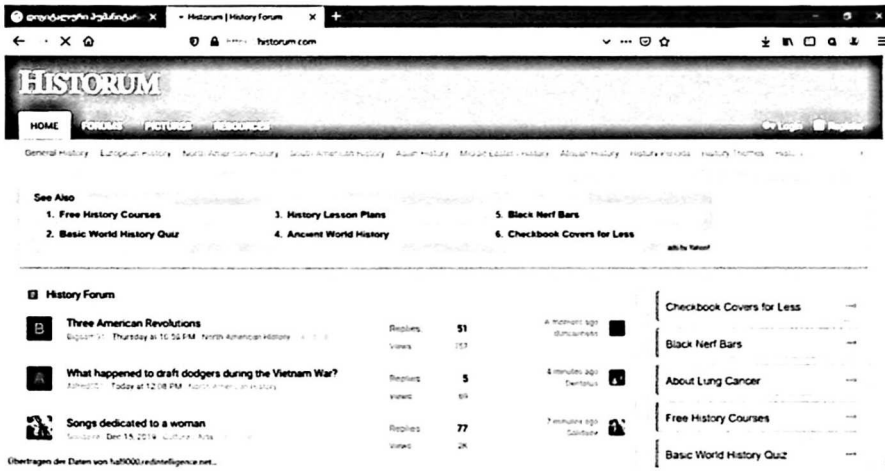
„Fancy, sexy und trendy"
Popkultur als Erinnerungspolitik: 60 Jahre Ungarn 1956

განსხვავებული კონცეფციისაა ე.წ. „ცოცხალი არქივის" პორტალი The September 11 Digital Archive, რომელიც ჯორჯ მაისონის უნივერსიტეტში (ვირჯინია) როი როზენცვაიგის სახელობის ისტორიისა და ახალი მედიების ცენტრში CHNM-ში (Roy Rosenzweig Center for History and New Media) შეიქმნა და მიზნად ისახავს 2001 წლის 11 სექტემბრის მოვლენების შესახებ მოგონებების (ცოცხალი ისტორიების) შეკრებას.[34]



[34] http://911digitalarchive.org

არსებობს თემატურად შეზღუდული პორტალები, როგორიცაა „ევროპული ისტორია ინტერნეტით“[35], რომელიც მაინცში, ლაიბნიცის ევროპული ისტორიის ინსტიტუტში შეიქმნა და ფორუმი historum.net, რომელიც ბავარიის სახელმწიფო ბიბლიოთეკის ელექტრონული გამოცემის ცენტრის ინტერნეტრესურსს წარმოადგენს. აღნიშნული პორტალები არ არიან თემატურად შეზღუდული და მხოლოდ თანამედროვე ისტორიაზე ფიქსირებული; ისინი მოიცავენ სხვადასხვა ტიპის რესურსებს საკვანძო თემებზე, რითაც ცდილობენ რაც შეიძლება ფართო აუდიტორიის მოზიდვას.



ბოლო პერიოდი აღინიშნება ახალი ტიპის პორტალების შექმნით – **თემატური პორტალების** გვერდით ჩნდება ე.წ. **ფუნქციური პორტალები**, როგორიცაა, მაგალითად:

1. „sehepunkte“, რომელიც ისტორიოგრაფიის დარგობრივ ელექტრონულ ჟურნალს წარმოადგენს, მხოლოდ ონლაინვერსიის სახით გამოდის და წელიწადში 12 ნომერს უშვებს;
2. „recensio“, რომელიც სხვადასხვა ჟურნალებიდან კრებს პუბლიკაციებს და ხელახლა აქვეყნებს ონლაინრეჟიმში[36];
3. დაბოლოს, სამეცნიერო ბლოგების პორტალები, რომელიც ასე მოზღვავდა ბოლო ხანს საინტერნეტო სივრცეში, უპირველეს ყოვლისა, ამერიკაში, საფრანგეთსა და ბოლო ხანს გერმანიაშიც. მაგალითად, ფრანგულენოვანი ბლოგპორტალი „hypoteses.org“, მოიცავს რამდენიმე ასეულ ბლოგს ჰუ-

[35] http://www.ieg-ego.eu

[36] იხ. http://www.sehepunkte.de და http://www.recensio.net

მანიტარულ და სოციალურ მეცნიერებათა დარგში. ბლოგპორტალს მიმდევრებიც გაუჩნდა – 2012 წლის დასაწყისში თანამოაზრეებმა გერმანიიდან, ავსტრიიდან და შვეიცარიიდან შექმნეს ბლოგპორტალის გერმანულენოვანი პროტოტიპი (spin-off) სახელწოდებით „de.hypoteses.org“;

4. ისტორიული ფოტოების პორტალი „Flickr Commons“[37], რომელიც აერთიანებს დაახლოებით ორმოცდაათამდე მეხსიერების ინსტიტუტს. მასში გაერთიანებულია, მაგალითად, დიდი ბრიტანეთის და აშშ-ს ეროვნული არქივი, სმიტსონის ინსტიტუტი და მრავალი კვლევითი და საბიბლიოთეკო დაწესებულება. პორტალის მიზანია ხელმისაწვდომი გახადოს მათ დაქვემდებარებაში არსებული ციფრული გამოსახულების კოლექციები;



5. ფუნქციურ პორტალებში ასევე ინტეგრირებულია „Docupedia-Zeitgeschichte“[38], რომელიც წარმოადგენს თანამედროვე ისტორიის კვლევის ცენტრალური ტერმინების, კონცეფციების, კვლევის მიმართულებებისა და მეთოდების ონლაინცნობარს.

ვებქსელის განვითარება, როგორც ვხედავთ, კვლავაც ხასიათდება სოციალური ინტერაქტიულობით, თუმცა შეინიშნება მისი ბლოგპორტალების გზით განვითარება.

[37] http://www.flickr.com/commons
[38] http://docupedia.de

2.4. დიგიტალური ჰუმანიტარიის განვითარების საკვანძო მოვლენები და რესურსები

დიგიტალური ჰუმანიტარიის განვითარების ისტორიაში უმნიშვნელოვანესი როლი შეასრულა სამეცნიერო დარგების კომპიუტერიზაციამ და ვებქსელის ტექნოლოგიურმა და კონცეპტუალურმა განვითარებამ. მისი დარგად ჩამოყალიბება მეტწილად განაპირობა დიდი მოცულობის დიგიტალური რესურსების შექმნამ, ერთი მხრივ, და აღნიშნული რესურსების კომპიუტერული დამუშავების ტექნოლოგიების განვითარებამ, მეორე მხრივ, რასაც რესურსებზე წვდომის შესაძლებლობები (ღია რესურსების კონცეფცია) დაემატა და რამაც გადამწყვეტი როლი შეასრულა დიგიტალური ჰუმანიტარიის, როგორც ინტერდისციპლინური (ტრანსდისციპლინური) დარგის ჩამოყალიბებაში. აქ, უპირველეს ყოვლისა, უნდა დავასახელოთ იტალიელი იეზუიტი მღვდელი **რობერტო ბუზა** (Roberto Busa), რომელმაც მისნად დაისახა დიდი მოცულობის მონაცემთა ბაზის შექმნა ტექსტების ინდექსაციის მისნით. ვინაიდან მის მიერ ჩაფიქრებულ ბაზაში საანალიზო სიტყვების რაოდენობა 11 მილიონს აჭარბებდა და მისი ხელით დამუშავება უდიდეს ადამიანურ რესურსებთან და დროსთან იყო დაკავშირებული, 1949 წელს რობერტო ბუსამ დახმარებისათვის მიმართა IBM-ის დამფუძნებელს **თომას უოტსონს** (Thomas J. Watson). დასმული ამოცანის გადასაჭრელად საჭირო პროგრამული უზრუნველყოფის შექმნის შედეგად დაიწყო საუკუნის მეგაპროექტის **Index Thomisticus** განხორციელება, რომელიც 30 წელს გაგრძელდა და რომლის შედეგები ბეჭდურად გამოცემულ 56 წიგნში არის ასახული. რობერტო ბუზას მიერ წამოწყებულმა პროექტმა **Index Thomisticus**, რომელიც მისნად ისახავდა თომა აქვინელის ნაშრომებისა და მისი მიმდევრების მიერ შექმნილი ტექსტების ინდექსაციას, საფუძველი დაუდო დიგიტალური ჰუმანიტარიის „დაბადებას".

გასული საუკუნის 60-იან წლებში იწყება კომპიუტერული ტექნოლოგიების აქტიური გამოყენება კონკორდანსის შესადგენად და ლექსიკოგრაფიული ამოცანების გადასაჭრელად. ამ პერიოდში იქმნება უმნიშვნელოვანესი რესურსები, რომლებმაც მნიშვნელოვანწილად განსაზღვრეს ჰუმანიტარული კომპიუტინგიდან დიგიტალურ ჰუმანიტარიაზე გადასვლის პროცესი. ქვემოთ ჩამოვთვლით რამდენიმე საკვანძო რესურსს და მოვლენას[39]:

---

[39] დიგიტალურ ჰუმანიტარიაში მნიშვნელოვანი პროექტების ჩამონათვალი იხ. ვებგვერდზე: https://blogs.brandeis.edu/library/2012/10/09/whats-digital-humanities-and-how-did-it-get-here/

1. 1966 წელს ტიუბინგენის უნივერსიტეტის მონაცემთა დამუშავების ცენტრში შეიქმნა ტექსტურ მონაცემთა დამუშავების პროგრამული პაკეტი **TUSTEP** (TUebinger System von TExtverarbeitungs-Programmen)[40]. პროგრამული პაკეტის შექმნის მიზანი იყო ტექსტების სამეცნიერო დამუშავების ავტომატიზაცია ჰუმანიტარულ დარგებში. პროექტის ფარგლებში შეიქმნა დიდი მოცულობის ტექსტურ მონაცემთა ბაზა, რომელიც ერთენოვანი პარალელური ტექსტების კომპიუტერული დამუშავების შედეგად ავტომატურად წარმოქმნიდა ტექსტის კრიტიკულ ვერსიასა და კრიტიკულ აპარატს. ამ მიმართულებით ერთ-ერთ წარმატებულ პროექტად შეგვიძლია დავასახელოთ იაკობ და ვილჰელმ გრიმების ლექსიკონის დიგიტალიზაცია. პროგრამული პაკეტი, რომელიც ვილჰელმ ოთმა (Wilhelm Ott) შექმნა, მუდმივად ვითარდება. 2018 წელს ექსპლუატაციაში შევიდა **TXSTEP** ვერსია – XML-ზე დაფუძნებული სინტაქსი TUSTEP-ისთვის. დღეისათვის პროექტის სამომხმარებლო პაკეტი გაფართოებულია: პროექტის ფარგლებში სისტემატურად ეწყობა კურსები, კოლოკვიუმები, კონფერენციები.
2. 1973 წელს დიგიტალურ ჰუმანიტარიაში მომუშავე მეცნიერები დარგის განვითარების მიზნით ერთიანდებიან და აარსებენ ლიტერატურული და ლინგვისტური კომპიუტინგის ასოციაციას (Association for Literary and Linguistic Computing), რომელიც საფუძველად დაედო 2011 წელს ინიცირებული ორგანიზაციის – **დიგიტალურ ჰუმანიტარიაში ევროპული ასოციაციის** The European Association for Digital Humanities (EADH) შექმნას. აღნიშნული ორგანიზაცია ამერიკულ ორგანიზაციასთან The Association for Computers in the Humanities (ACH) და კანადურ ორგანიზაციასთან Society for Digital Humanitics / Société pour l'étude des médias interactifs (SDH-SEMI) ერთად 2005 წელს აარსებს **საერთაშორისო ორგანიზაციების ალიანსს დიგიტალურ ჰუმანიტარიაში** The Alliance of Digital Humanities Organizations (ADHO). დღეისათვის ალიანსში გაერთიანებული არიან **ავსტრალიური ასოციაცია** Australasian Association for Digital Humanities (aaDH)[41], **იაპონური ასოციაცია** Japanese Association for Digital Humanities (JADH)[42], **ფრანგულენოვანი ასოციაცია დიგიტალურ ჰუმანიტარიაში** (The french-speaking association for Digital Humanities. ასევე centerNet[43] და Humanistica[44].

[40] http://www.tustep.uni-tuebingen.de/
[41] https://aa-dh.org/
[42] https://www.jadh.org/
[43] http://dhcenternet.org/
[44] http://www.humanisti.ca/

3. 1985 წელს ჰარვარდის უნივერსიტეტში საფუძველი ეყრება მასშტაბურ პროექტს ჰუმანიტარულ დარგში – **პერსევსის დიგიტალურ ბიბლიოთეკას** (Perseus Digital Library)[45]. დიგიტალური ბიბლიოთეკის ძირითად და მნიშვნელოვან ნაწილს შეადგენს ანტიკური პერიოდის წყაროები ბერძნულ და ლათინურ ენებზე, თუმცა ბიბლიოთეკის თემატიკა იწრდება და დღეისათვის, მაგალითად, ამერიკის ისტორიის მე-19 საუკუნის ამსახველ მასალებსაც მოიცავს.
4. 1986 წელს პირველად გამოდის სამეცნიერო პროფესიული ჟურნალი დიგიტალურ ჰუმანიტარიაში „**ლიტერატურული და ლინგვისტური კომპიუტინგი**" Literary and Linguistic Computing (LLC).
5. 1987 წელს დაარსდა **ტექსტის კოდირების ასოციაცია** Text Encoding Initiative (TEI), რომელმაც შეიმუშავა კოდირების საერთაშორისო სტანდარტი რესურსების სისტემური გადამუშავებისა და მონაცემთა და მეტამონაცემთა გაცვლის მიზნით. 1988 წელს შექმნილი და 1990 წელს გამოქვეყნებული ვერსია P1 (proposal 1) ოთხი წლის შემდეგ შეიცვალა P2 და P3 ვერსიებით. 2000 წელს, მონაცემთა დამუშავების პროცესში კოდირების საერთაშორისო დონეზე სტანდარტიზაციის აუცილებლობიდან გამომდინარე, დაარსდა კონსორციუმი (TEI-Consortium), რომელმაც 2002 წელს P4 და 2005 წელს P5 ვერსიები განავითარა. ეს უკანასკნელი ვერსია საპილოტო ფაზის გაელის შემდეგ (2007 წლიდან) საერთაშორისო მიმოქცევაშია და წინა ვერსიებისაგან განსხვავებით ხელნაწერთა აღწერის თანამედროვე სტანდარტსაც გვთავაზობს.
6. 1998 წელს პირველად გაიცა **სამეცნიერო პრემია დიგიტალური ჰუმანიტარიის განვითარებაში განსაკუთრებული წვლილის აღსანიშნავად**, რომელიც **რობერტო ბუზას** სახელს ატარებს.
7. 2004 წელს დაიწყო მასშტაბური პროექტი **Google Books**, რომელიც მიზნად ისახავდა (როგორც პირველადი, ისე მეორადი) სამეცნიერო რესურსების საერთაშორისო ხელმისაწვდომობას და რომელმაც დიდად შეუწყო ხელი **ღია რესურსების** იდეის განვითარებას.
8. 2010 წელს შეიქმნა **Google Ngram**, რომელიც გამოიყენება ქსელში სოციალური და კულტურული მიმართულებების საანალიზოდ.
9. 2011 წელს ჟურნალი „მეცნიერება" (Science) აქვეყნებს **ელერ ლიბერმან აიდენისა** (Erez Lieberman Aiden) და **ჟან-ბატისტ მიშელის** (Jean-Baptiste Michel) სტატიას **კულტურომიკის** საკითხებზე „კულტურის რაოდენობრივი ანალიზი მი-

[45] http://www.perseus.tufts.edu/hopper/

ლიონობით დიგიტალური წიგნის გამოყენებით“ (Quantitative Analysis of Culture Using Millions of Digitized Books), რომელმაც გადატრიალება მოახდინა დიგიტალურ ჰუმანიტარიაში – საერთაშორისო მიმოქცევაში შევიდა ნეოლოგიზმი „კულტურომიკი“. სტატიაში პირველად არის განსაზღვრული **კულტურომიკის** ცნება და მისი აღმნიშვნელი ტერმინი. ავტორები ამტკიცებდნენ, რომ დიდ მონაცემთა ბაზების სტატისტიკური დამუშავების შედეგად შესაძლებელია არა მარტო ადამიანთა ქცევის უკეთ გაგება, არამედ პროგნოზირებაც კი. **კულტურომიკი 2.0** ვერსიის ფარგლებში **ქალევ ლეტარუმ** (Kalev H. Leetaru)[46] გამოიკელია **ახალი ამბების** ტექსტები (როგორც ბეჭდვითი, ისე მედიარესურსები) და დაადგინა, რომ კულტურომიკის მეთოდური ინვენტარის გამოყენებით, მაგალითად. შესაძლებელი იყო არაბული გაზაფხულის (იგივე პანარაბული რევოლუციის) პროგნოზირება, როდესაც ტუნისში დაწყებული საპროტესტო გამოსვლები სულ მალე გადაიქცა მასობრივ-სახალხო მოძრაობად არაბულ სამყაროში, ჩრდილოეთ აფრიკასა და ახლო აღმოსავლეთში და რომელსაც 60000-ზე მეტი ადამიანის სიცოცხლე შეეწირა მსხვერპლად. ხოლო დიდი მოცულობის დიგიტალურ მონაცემთა დამუშავების შედეგად აღმოჩენილი კანონზომიერება, რომელიც განწყობის გამომხატველი სიტყვების კონკრეტულ გეოგრაფიულ არეალში მათი სიხშირული გამოყენების საფუძველზე დადგინდა, საფუძვლად დაედო ოსამა ბინ ლადენის ადგილსამყოფლის განსაზღვრას 200 კილომეტრი რადიუსის სიზუსტით. 2012 წელს **ალექსანდრე პეტერსენისა** (Alexander M. Petersen) და მისი კოლეგების მიერ აღმოჩენილ იქნა ენაში სიტყვების დაბადებისა და სიკვდილის, ანუ მათი არსებობის ხანგრძლივობის კანონზომიერება. კერძოდ, აღმოჩნდა, რომ სიტყვათა „კვდომის“ შემთხვევების რაოდენობა გაცილებით მაღალია, ვიდრე სიტყვათა „დაბადების“. შესაბამისად, დადგინდა სიტყვათა „სიცოცხლისუნარიანობა“, რაც საფუძვლად დაედო ლექსიკონებში სალექსიკონო ერთეულთა სელექციის (სიტყვათა შეტანა-არშეტანის) პრინციპს[47]. **კულტურომიკი** საშუალებას იძლევა სოციალური მედიების რესურსებზე დაყრდნობით ვიკვლიოთ საზოგადოებრივი აზრის დინამიკა. სხვათა შორის, 27 ქვეყნის 1,3 მილიონი სტატიის დამუშავების შედეგად აღმოჩნდა, რომ თემები,

[46] Kalev H. Leetaru, Culturomics 2.0: *Foresasting lange-scale human behavion using global news media tone in time and space*. https://firstmonday.org/article/view/3663/3040

[47] Alexander M. Petersen, Joel Tenenbaum, Shlomo Havlin & H. Eugene Stanley, *Statistical Laws Governing Fluctuations in Word Use from Word Birth to Word Death,*. Scientific Reports. 2: 313. (15 March 2012) https://www.nature.com/articles/srep00313

რომელსედაც ამახვილებენ ყურადღებას ამა თუ იმ ქვეყნის მედიები, მსგავსია გეოგრაფიულ, კულტურულ და ეკონომიკურად მსგავს ქვეყნებში. იგივე შედეგი მიიღეს მეცნიერებმა ევროპიის შეფასების შემთხვევაში: მაყურებლის სიმპატიების მსგავსება ქვეყნის გარკვეული პარამეტრების მიხედვით მსგავსების ადეკვატურია.

2.5. მონაცემთა აფეთქების საუკუნე

დიგიტალური ჰუმანიტარიის სამეცნიერო ინტერდისციპლინურ დარგად ჩამოყალიბების ერთ-ერთ მნიშვნელოვან ასპექტს წარმოადგენს ისტორიულად მნიშვნელოვანი რესურსების ფართომასშტაბიანი გაციფრება – ე.წ. **რეტროდიგიტალიზაცია**. ამ პროცესის შედეგად დღეს მანქანურად წაკითხვადია ისტორიული რესურსების მონაცემთა უსარმაზარი რაოდენობა. გაციფრებული რესურსები აღჭურვილია მეტამონაცემებით, რაც მათი ეფექტიანი გამოყენების საშუალებას იძლევა. ნებისმიერ შემთხვევაში, წყაროების ის მოცულობა, რომელიც დღეს ინტერნეტის საშუალებით წვდომადია, წარმოუდგენელი იყო რამდენიმე წლის წინ. დიგიტალიზირებული მასალის მნიშვნელოვანი ნაწილი მოდის პროექტის Google Books ფარგლებში შექმნილ რესურსებზე, რომელიც მიზნად ისახავდა 85-დან 130-მდე მილიონი წიგნის გაციფრებას და მის ინტერნეტში განთავსებას.

ამ დიდი რაოდენობით მონაცემებთან დაკავშირებით ჩნდება კითხვა, შესაძლებელია თუ არა თანამედროვე რესურსების ახალი, კომპიუტერული მეთოდებით დამუშავება, რამდენად სანდოა მიღებული შედეგი და რამდენად ადეკვატურია კვლევის შედეგები. აქ უპირველეს ყოვლისა საუბარია მეთოდოლოგიურ მიდგომებზე და იმ დასკვნებზე, რომლებიც შეიძლება მივიღოთ ამგვარი მონაცემების ანალიზის შედეგად. „მონაცემთა ამოძრავების ისტორია“ კრიტიკის მთავარი პუნქტი წყაროების საკითხს უკავშირდება. ამის საილუსტრაციოდ შეიძლება მოვიყვანოთ ყველაზე ცნობილი პროექტის მაგალითი: ონლაინინსტრუმენტს „Ngram Viewer“ შეუძლია გააანალიზოს ასოთა თანმიმდევრობა 5.2 მილიონი წიგნის ბაზაზე. მარტივად აგებული საძიებო ველი საშუალებას გაძლევთ შეიყვანოთ საძიებო სიტყვა ან ტერმინი, მიუთითოთ ძიების პროცესისათვის რელევანტური ენა და ძიების ქრონოლოგიური ჩარჩო. ძიების შედეგად თქვენ მიიღებთ გრაფიკულად ვიზუალიზებულ დიაგრამას, რომელიც ასახავს წიგნების რაოდენობას, რომლებშიც ერთხელ მაინც იძებნება ასოთა აღნიშნული თანმიმდევრობა. მიღებული შედეგი გვიჩვენებს ასევე დასკანერებული წიგნების რაოდენობის პროცენტულ მაჩვენებელს მოცემულ ქრონოლოგიურ ჩარჩოში. ეს ბა-

დებს კითხვას, არის თუ არა ამგვარი ინსტრუმენტი „სახალისო გართობა“ თუ კვლევის სტატისტიკური მეთოდი, რომელიც თვისებრივი დასკვნების პროგნოზირების საშუალებას გვაძლევს? რამდენად სანდოა ეს დასკვნები? ამგვარი ძიების შედეგად მივიღეთ კი ის, რის მიღებასაც სხვაგვარად ვერ შევძლებდით? ნებისმიერ შემთხვევაში, ძიების აქ აღწერილი ინსტრუმენტი შესაძლებელს ხდის დიდი მოცულობის ტექსტურ კორპუსს კონკრეტული შეკითხვები დაუსვათ კონკრეტულ ფაქტებთან მიმართებაში. ამრიგად, ეს არის ინსტრუმენტი, რომელსაც შეუძლია **სტატისტიკური კვლევის შედეგები** აქციოს **კვლევის ამოცანებად**. სტატისტიკური მეთოდის გამოყენების პრაქტიკულ მნიშვნელობას ქვემოთ საგანგებოდ შევეხებით ქვეთავში „სტატისტიკა ლინგვისტიკის სამსახურში“.

Ngram Viewer შეიძლება გაიგოთ, როგორც ექსპერიმენტული ლაბორატორია, რომელიც ამუშავებს დიდ ტექსტურ ბაზებს მათი ჰუმანიტარული და კულტუროლოგიური ასპექტით კვლევის მიზნით. დიგიტალურ ჰუმანიტარიაში გამოყენებული ტექსტური კორპუსის რესურსები თარიღდება წიგნის ბეჭდვის დასაწყისიდან დღემდე. საუკეთესო შედეგები დღემდე მიღწეულია ინგლისურენოვანი ტექსტებისთვის, რომლებიც 1800-დან მოყოლებული 2000 წლებამდე პერიოდში შეიქმნა.

ისტორიული კვლევებისათვის Ngram Viewer-ის (ან მსგავსი პრინციპით შექმნილი პროდუქტების) გამოყენება შეიძლება ძალიან მნიშვნელოვანი აღმოჩნდეს, განსაკუთრებით ისტორიული დისკურსის რაოდენობრივი ანალიზის შემთხვევაში, თუ მას დავუმატებთ ტერმინის ისტორიას და ისტორიულ სემანტიკას.

ეს ამოცანა მეთოდურად დაკავშირებულია ისეთი დიდი ციფრული ტექსტური რესურსების გამოყენებასთან, როგორიცაა, მაგალითად, გერმანული ენის დიგიტალური ლექსიკონი Digitales Wörterbuch der deutschen Sprache (DWDS)[48], რომელიც აერთიანებს გერმანული ენის ლექსიკას 1600 წლიდან მოყოლებული დღემდე და მოიცავს სხვადასხვა ლექსიკონებში დადასტურებულ 585 000 სალექსიკონო ერთეულს, 27 მილიარდ კონტექსტს ისტორიული და თანამედროვე ტექსტებიდან, რაც სიტყვათა სიხშირისა და შესიტყვებების დიაქრონიული ანალიზის საშუალებას იძლევა 400 წლის ფარგლებში. აღნიშნული რესურსი მუდმივად განახლებადია. მაგ, პანდემიის დაწყებისთანავე შეიქმნა ახალი თემატური ლექსიკონი DWDS-Themenglossar zur COVID-19-Pandemie[49], რომელიც გერმანულ ენაში შემოსულ და კოვიდთან დაკავშირებით გამოყენებულ ახალ სიტყვებს და შესიტყვებებს აღრიცხავს და აანალიზებს.

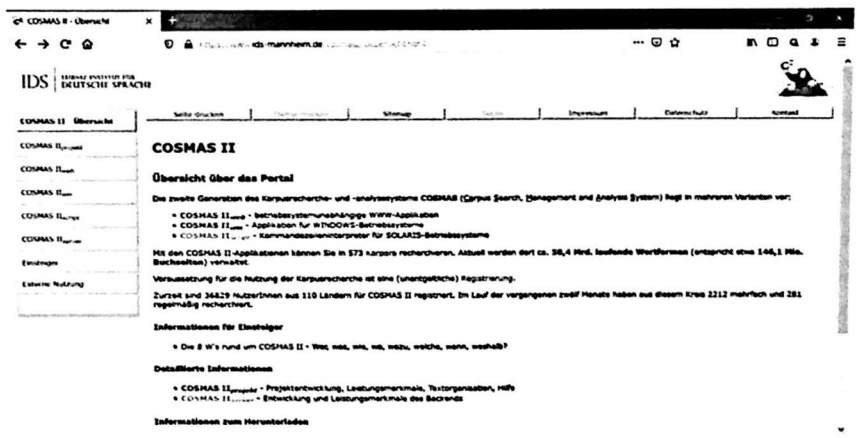
[48] https://www.dwds.de/

[49] https://www.dwds.de/themenglossar/Corona



DWDS-ის ვებგვერდი

გერმანული ენისათვის შექმნილ დიდ მონაცემთა შორის უნდა დავასახელოთ ასევე მეგაკორპუსი Corpus Search, Management and Analysis System (COSMAS II).[50] აღნიშნული კორპუსის მომხმარებელს წვდომა აქვს 573 კორპუსზე, რომლის საერთო მოცულობა 58,4 მილიარდ ტოკენს შეადგენს. აქტუალური მონაცემებით 110 ქვეყნის დაახლოებით 37000-მდე მომხმარებელი სარგებლობს COSMAS II-ის აპლიკაციით.



COSMAS II-ის ვებგვერდი

ისტორიულ მეცნიერებაში დიდი ოდენობით მონაცემთა დამუშავების საკითხთან მჭიდრო კავშირშია შედეგების ციფრულად წარმოდგენის შესაძლო ფორმების

[50] https://www.ids-mannheim.de/cosmas2/

საკითხი. Ngram Viewer არ წარმოადგენს რიცხვების გროვას, არამედ შედეგების (თუმცა ძალიან მარტივ) ვიზუალიზაციას. ისტორიული კონტექსტების ვიზუალიზაცია წარმოადგენს ახალ შესაძლებლობას, რომლის პოტენციალი ჯერ ბოლომდე არ არის გამოყენებული. ამ სფეროში აპლიკაციების უმეტესობა ყურადღებას ამახვილებს იმ ისტორიული რუქების გამოყენებაზე, რომლებზეც მონაცემები განთავსებულია როგორც დამატებითი ინფორმაცია. განვითარების საოცარი დინამიკით გამოირჩევა გეოგრაფიული ინფორმაციული სისტემები (GIS), რომელიც ჩვენ თვალწინ ჩამოყალიბდა ახალ სამეცნიერო დარგად[51].

ვიზუალიზაციის ახალი, მაგალითად, ინტერაქტიულად ფუნქციონირებადი ხელსაწყოების შექმნასა და გამოყენებას უდიდესი მნიშვნელობა ენიჭება როგორც ისტორიის ან გეოგრაფიის მეცნიერებისათვის, ისე საზოგადოებრივი ისტორიის გაგების და მისი ფუნქციონირების თვალსაზრისით.

2.6. ღია რესურსები და მათი მნიშვნელობა

დიგიტალური ჰუმანიტარიის სწრაფი განვითარება ძირითადად **დიდ მონაცემთა ბაზების (big data)** შექმნით იყო განპირობებული, თუმცა ერთ-ერთ გადამწყვეტ ფაქტორს ე.წ. **ღია რესურსების კონცეფციის** შექმნა წარმოადგენდა, რომელმაც რესურსების სამეცნიერო მიმოქცევის სრულიად ახალი დინამიკა შექმნა.

მეცნიერული კომუნიკაციის ახალი ფორმების ევოლუციამ (ვებტექნოლოგიების განვითარებამ) უპრეცედენტო შესაძლებლობები შეუქმნა მეცნიერებს პოტენციური გლობალური კავშირების მეშვეობით სწრაფად და ეფექტიანად გამოეყენებინათ ელექტრონულ საცავებში დაცული დიგიტალური რესურსები. რესურსებზე წვდომის შესაძლებლობა სულ მალე გადაიქცა **ღია რესურსების კონცეფციად**, რამაც თავის მხრივ დიგიტალური არქივების ინფრასტრუქტურის განვითარებას შეუწყო ხელი და თვისებრივად შეცვალა 21-ე საუკუნის ჰუმანიტარიაში კვლევის პროცესი. მონაცემთა დიდ ბაზებზე თანდართული საძიებო სისტემები საშუალებას იძლევა მიზნობრივად ვეძებოთ და მოვიპოვოთ ჩვენთვის საინტერესო მასალა. ამისათვის საკმარისია საძიებო რესურსების კონკრეტული პარამეტრების განსაზღვრა, და საძიებო სისტემა თავად მოახდენს საჭირო მასალის სელექციას. ამასთან, დიდ მონაცემთა ბაზები უხილავია მომხმარებლისთვის და, შესაბამისად, დაცულიც, ვინაიდან შესღუდულია მათზე სრული ხელმისაწვდომობა და რაც მთავარია, მათი კოპირება-გავრცელება. რესურსების ე.წ. „გადაჭრის“ აუცილებლობა დღეისათვის არც არსებობს, ვინაიდან

[51]https://www.esri.com/en-us/what-is-gis/history-of-gis

ენობრივი რესურსების საცავების ღია სისტემად გარდაქმნა და რესურსების მოპოვების საყოველთაობა თავად გამორიცხავს რესურსების „მისაკუთრების“ მოტივაციას. პირიქით, მეცნიერებს საშუალება მიეცათ საკუთარი რესურსების გახსნისა და გაერთიანების გზით – ერთიანი „მსოფლიო ბიბლიოთეკის“ შექმნით – შეემუშავებინათ საერთაშორისო სამეცნიერო კომუნიკაციის ახალი ფორმები და განევითარებინათ რესურსების მოპოვებისა და კვლევაში მათი ეფექტიანი გამოყენების ახალი მეთოდები.

„ჩვენ მოვუწოდებთ მეცნიერებს, მონაწილეობა მიიღონ ღია ვებარქივების შექმნაში, რათა ღირსეულად ვუპასუხოთ აკადემიური ბიბლიოთეკების წინაშე წარმოშობილ პრობლემებსა და გამოწვევებს, ჩვენ ერთად უნდა შევქმნათ არქივების თანამედროვე ფორმები და მეთოდები, გავხდეთ რესურსების აკადემიური მოხმარების შუამავლები და ხელი შევუწყოთ რესურსების დაარქივების, ორგანიზებისა და შენახვის საერთაშორისო სტანდარტების შემუშავებას მათი საყოველთაო მოხმარების მიზნით“ (თ.) – წერდა იუნგ-რან პარკი (Jung-ran Park) ჟურნალ Electronic Journal of Academic and Special Librarianship-ში გამოქვეყნებულ სტატიაში „Language-Related Open Archives: Impact on Scholarly Communities and Academic Librarianship“[52].

ვებტექნოლოგიის მარკეტინგის განვითარებამ, რომელიც სრულიად განსხვავდება ტრადიციული, ე.წ. „ქაღალდის ეპოქის“ მარკეტინგისაგან, უპრეცედენტო შესაძლებლობები მოუტანა სამეცნიერო საზოგადოებას – ინფორმაციის ინტერნეტით გავრცელება გაცილებით მეტი სისწრაფით გამოირჩევა, ინტერნეტტექნოლოგიები და ინფრასტრუქტურა კი – მეტი მოქნილობით. სამეცნიერო კომუნიკაციის ამ მაღალეფექტიანმა საშუალებებმა, ერთი მხრივ, და ქსელური კომუნიკაციის ახალი ფორმების განვითარებამ, მეორე მხრივ, აუცილებლობით გამოიწვია ტრადიციული სამეცნიერო კომუნიკაციის არსებული საშუალებების შეზღუდვა. სამეცნიერო კომუნიკაციის აქამდე არსებული ფორმები მიუღებელი აღმოჩნდა თანამედროვე საზოგადოებისათვის ეფექტიანობის თვალსაზრისით. კერძოდ:

1. სამეცნიერო კომუნიკაციის ტრადიციული ფორმები გაცილებით მოუქნელია მზარდი ტექნოლოგიური განვითარების პირობებში როგორც სამეცნიერო გავრცელების სისწრაფის, ისე ფინანსური თვალსაზრისით. სამეცნიერო კომუნიკაციის ტრადიციულ, ანუ ბეჭდურ ფორმებს ახასიათებს წარმოების შეუდარებლად მაღალი ხარჯები, ვიდრე კომუნიკაციის (ანუ გავრცელების) ელექტრონულ საშუალებებს;

---

[52] Jung-ran Park, *Language-Related Open Archives: Impact on Scholarly Communities and Academic Librarianship*, (http://southernlibrarianship.icaap.org/content/v05n02/park_j01.htm)

2. სამეცნიერო ნაშრომების შენახვა ლოგისტიკურ და ფინანსურ ხარჯებთან არის დაკავშირებული: სტამბის საწყობებიდან დაწყებული დამთავრებული აკადემიური სივრცით (ბიბლიოთეკა, არქივი);
3. ეს უკანასკნელი მოითხოვს ადამიანურ რესურსებს, აღჭურვილს საგანგებო სამომხმარებლო ცოდნით;
4. ბეჭდური რესურსების გავრცელება გაცილებით შეზღუდულია როგორც რაოდენობრივად (ტირაჟირება), ისე გეოგრაფიულად, რაც, შესაბამისად, კიდევ უფრო გართულებულია რესურსების მომხმარებლამდე მიტანის გაშუალებული პროცედურით (რესურსების შექმნა/წარმოებიდან მათ გავრცელებამდე საჭირო დრო ბევრად აღემატება ელექტრონულად გამოქვეყნების შემთხვევაში მომხმარებლამდე მისი მიღწევის დროს).

დღეს ეჭვს აღარ იწვევს ის ფაქტი, რომ **მასობრივი შენახვის ტექნოლოგიები**, რომელიც **ტრადიციული შენახვის** ფორმებთან შედარებით გაცილებით იაფია, საშუალებას იძლევა განუსაზღვრელი რაოდენობის რესურსი შეინახოთ ციფრულ ფორმატში. გარდა ამისა, ინტერნეტკომუნიკაცია საშუალებას იძლევა ტრადიციული სამეცნიერო **პუბლიკაციის სტატიკური ფორმები** (ტექსტური ფორმატი) გარდაექმნათ **დინამიკურ მულტიმედიალურ** და **მულტიმოდალურ რესურსად** და გავამდიდროთ იგი, მაგალითად, ინტეგრირებული ხმით, მოდიფიცირებული ტექსტებით, ტრანსკრიპტებით, ვიზუალური გამოსახულებებით, ლინკებით და ა.შ. ამგვარად წარმოდგენილი რესურსები, ბუნებრივია, სამეცნიერო კომუნიკაციის მაღალი ხარისხით გამოირჩევა.

მასობრივი შენახვის დიგიტალურმა ტექნოლოგიებმა მრავალჯერადი გამოყენების გახადა სამეცნიერო რესურსები და საფუძველი დაუდო **„დიგიტალურ აღწარმოებას“**; მაგალითად, გააადვილა ტექსტების **ინდექსირება** და გლოსირება, რესურსების **კონვერტირება** და ახალ დიგიტალურ ფორმატში მათი გადაყვანა, უსრულყო ფართომასშტაბიანი მიზნობრივი **რეპროდუქცია** და **სტატისტიკური დამუშავება**, და რასაკვირველია, შეუდარებლად გაზარდა სამეცნიერო თემებისა და პრობლემების საზოგადოებისათვის **ხელმისაწვდომობა**. ამ თვალსაზრისით, ახალი, ციფრული წარმოების ფორმები და რესურსების გავრცელების შეუსღუდავი საშუალებები გლობალურ ქსელში გავრცელების ფართო პოტენციალით ხასიათდება.

გლობალურ საკომუნიკაციო სისტემაში დიგიტალური რესურსების შეუფერხებელი მიმოქცევა, ბუნებრივია, **სტანდარტიზაციის**, განსაკუთრებით კი **მეტამონაცემების სტანდარტიზაციის** პრობლემის წარმატებით გადაჭრას უკავშირდება. მონაცემთა კლასიფიკაციის, რესურსების ორგანიზაციისა და სტანდარტიზებული ინდექსაციის სქემის გარეშე, საძიებო მექანიზმის ეფექტიანობა, და შესაბამისად, რესურსების

მოპოვების სიზუსტის ხარისხი საგრძნობლად დაბლდება. გლობალური საკომუნიკაციო სისტემის ფუნქციონირების ასევე მნიშვნელოვან წინაპირობას, მზარდი ინტერნეტრესურსების საცავების შექმნის პირობებში, წარმოადგენს **თავსებადი ტექნოლოგიური ინფრასტრუქტურის** შექმნა.

გლობალური საკომუნიკაციო სისტემის ფუნქციონირების პროცესში წარმოქმნილმა პრობლემებმა და გამოწვევებმა, როგორიცაა მაგ.: ა) ძიების შედეგების უზუსტობა და ძიების პროცესის არაეფექტიანობა (განმეორებითი ძიების სიხშირე), ბ) ინდექსაციის ტერმინების სტანდარტიზაციის აუცილებლობა, გ) დიგიტალური მონაცემების სხვადასხვა ფორმატში შენახვის პრაქტიკა, დ) მონაცემთა განსხვავებული კოდირების სქემების გამოყენება, ე) ლინგვისტური და კულტუროლოგიური მონაცემების მოპოვების, დაარქივებისა და შენახვის სტანდარტიზაციის აუცილებლობა, და იმის გაცნობიერებამ, რომ მზარდი ვებტექნოლოგიების პირობებში გლობალურ საკომუნიკაციო სისტემას გაცილებით მეტი პოტენციალი გააჩნია, განაპირობა სამი უმნიშვნელოვანესი ინიციატივის შექმნა. კერძოდ:

- **დუბლინის ბირთვის მეტამონაცემების ინიციატივა** Dublin Core Metadata Initiative **(DCMI)**, რომელიც 1995 წელს შეიქმნა დუბლინში (Dublin, Ohio)[53];
- **ღია არქივების ინიციატივა** Open Archives Initiative **(OAI)**, რომელიც 1999 წელს შეიქმნა სანტა ფეში (Santa Fe)[54];
- **ენობრივი რესურსების ღია არქივების საზოგადოება** Open Language Archives Community **(OLAC)**, რომელიც 2000 წელს შეიქმნა ფილადელფიაში (Philadelphia)[55].

დიგიტალური ჰუმანიტარიის განვითარება, როგორც ზემოთ უკვე ვიხილეთ, არაერთხაზოვანი იყო, სხვადასხვა მიმართულებებსა და ასპექტებს მოიცავდა და ჯერ კიდევ ფორმირების პროცესში იმყოფება, არ დასრულებულა. დიგიტალური ჰუმანიტარიის ჩამოყალიბება, რომელიც სამეცნიერო პროექტების ტექნოლოგიზების განვითარების აუცილებლობით იყო განპირობებული, განვითარების რამდენიმე ეტაპს მოიცავს:

1. დიგიტალური ჰუმანიტარიის განვითარების პირველი ეტაპი (Digital Humanities), რომელიც ძირითადად ტექსტების დიგიტალიზაციისა და ტექსტების ელექტრონული დამუშავების ტექნოლოგიების განვითარებით ხასიათდება;

---

[53] Dublin Core Metadata Initiative: http://dublincore.org (October 30, 2004)
[54] Open Archives Initiative: http://www.openarchives.org (October 30, 2004).
[55] Open Language Archives Community: http://www.language-archives.org (October 30, 2004)

2. დიგიტალური ჰუმანიტარიის განვითარების მეორე ეტაპი (Digital Humanities 2.0), რომელიც ხასიათდება სამეცნიერო დისციპლინების სინერგეტიკით, კონვერგენტული და ჰიბრიდული სფეროების შემოტანით, და რაც უაღრესად მნიშვნელოვანია, გენერატიულობით. ამ ეტაპზე იქმნება და იხვეწება დიდ მონაცემთა შექმნის, მოხმარებისა და შენახვა-დაარქივების ძირითადი პრინციპები და მეთოდები, საფუძველი ეყრება გამოქვეყნების ინოვაციურ ფორმებსა და საშუალებებს, რომელიც პრინციპულად განსხვავდება ტრადიციული საგამომცემლო კონცეფციისაგან.

როგორც ჯეფრი შნაპი (Jeffrey Schnapp), პეტერ ლუნენფელდი (Peter Lunenfeld) და თოდ პრესნერი (Todd Presner) მიუთითებენ დიგიტალური ჰუმანიტარიის მანიფესტში (Digital Humanities Manifesto 2.0, 2009), ციფრული ჰუმანიტარული მეცნიერების ჩამოყალიბების პირველი ეტაპი რაოდენობრიობაზე (მონაცემთა ბაზის შექმნა, მოძიების შესაძლებლობებისა დახვეწა და საძიებო სისტემების განვითარება) და კორპუსლინგვისტიკის განვითარებაზე იყო ორიენტირებული, მეორე ეტაპი კი გამოირჩევა თვისობრიობით, გენერატიულობით, ინტერპრეტაციულობით, ორიენტირებულია გამოცდილებაზე და ემოციურობით ხასიათდება. დიგიტალური ჰუმანიტარია ავითარებს და იყენებს კვლევის ციფრულ ინსტრუმენტებს, რითაც მყარ ტექნოლოგიურ საფუძველს უქმნის თანამედროვე ჰუმანიტარულ კვლევებს და საგრძნობლად ამაღლებს კვლევის ხარისხს, რომელიც დიგიტალური ჰუმანიტარიის ფარგლებში ხასიათდება მეთოდოლოგიური სისტემურობით, კომპლექსურობით, მედიალური სპეციფიკურობით, ისტორიული კონტექსტით, ანალიტიკურობის სიღრმით, კრიტიკისა და ინტერპრეტაციის გათვალისწინებით[56].

დიგიტალური ჰუმანიტარიის განვითარების მეორე ეტაპი იმითაც გამოირჩევა, რომ იქმნება დარგობრივი საერთაშორისო ორგანიზაციები და ვითარდება დარგობრივი საგამომცემლო ინფრასტრუქტურა. კერძოდ:

საერთაშორისო ორგანიზაციები:

- *The Association for Computers in the Humanities* (ACH),
- *European Association for Digital Humanities* (EADH) (2011 წლამდე *Association for Literary and Linguistic Computing* (ALLC))
- *Society for Digital Humanities / Société pour l'étude des médias interactifs* (SDH-SEMI)
- *The Alliance of Digital Humanities Organizations* (ADHO)

[56] იხ. https://www.humanitiesblast.com/manifesto/Manifesto_V2.pdf

პერიოდული გამოცემები დიგიტალურ ჰუმანიტარიაში:

- *Literary and Linguistic Computing,* 1986 წლიდან
- *Forum Computerphilologie,* 1999 წლიდან
- Computers and the Humanities, 1966 წლიდან; 2005 წლიდან კი გამოიცემა სახელწოდებით *Language Resources and Evaluation*
- Literary and Linguistic Computing (LLC), 1986 წლიდან; 2015 კი გამოიცემა სახელწოდებით *Digital Scholarship in the Humanities* (DSH)
- Digital Studies / Le champ numérique, 1992 წლიდან
- Journal of Digital Information, 1997 წლიდან
- Forum Computerphilologie, 1999–2012 წლები
- Digital Medievalist, 2005 წლიდან
- International Journal of Digital Curation, 2006 წლიდან
- Digital Humanities Quarterly (DHQ), 2007 წლიდან
- digiversity – Webmagazin für Informationstechnologie in den Geisteswissenschaften, 2011 წლიდან
- Journal of Digital Humanities (JDH), 2011–2014 წლები
- Caracteres – Estudios culturales y criticos de la esfera digital, 2012 წლიდან
- Zeitschrift für digitale Geschichtswissenschaften, 2012–2015 წლები
- Digital Philology: A Journal of Medieval Cultures, 2012 წლიდან
- RIDE – A review journal for digital editions and resources, 2014 წლიდან
- Digital Classics Online, 2015 წლიდან
- Zeitschrift für digitale Geisteswissenschaften, 2015 წლიდან
- Journal of Cultural Analytics, 2016 წლიდან

დიგიტალური ჰუმანიტარია, როგორც მეცნიერების ნებისმიერი სფერო, კვლევის ობიექტის განსაზღვრის, ძირითადი პრინციპების შემუშავებისა და კვლევის მეთოდების დადგენისა და დახვეწის გზით ვითარდება. დიგიტალური ჰუმანიტარიის ქართველოლოგიაში შემოტანამ ქართველოლოგიის ტრანსფორმაცია განაპირობა და დიგიტალური ქართველოლოგიის წარმოშობას შეუწყო ხელი.

თავი მეორე
დიგიტალური ქართველოლოგია

1. დიგიტალური ქართველოლოგიის განვითარების ისტორია, ეტაპები და ტექნოლოგიები

დიგიტალური ქართველოლოგია დიგიტალური ჰუმანიტარიის პარადიგმის ერთ-ერთ განშტოებას წარმოადგენს და ყველა იმ ნიშნით ხასიათდება, რომელიც ამ უკანასკნელს გააჩნია. თუ ცნებას „ჰუმანიტარია“ დავაეიწროებთ და დავიყვანთ ცნებამდე „ქართველოლოგია“, მაშინ ცნება „დიგიტალური ქართველოლოგია“ შეიძლება განვმარტოთ, როგორც ენათმეცნიერების ერთი კონკრეტული უბანი, რომელიც კვლევის პროცესში იყენებს გაციფრებული რესურსებისაგან შემდგარ დიდ მონაცემთა ბაზებს, მონაცემთა სისტემური დამუშავების საგანგებო პროგრამულ საშუალებებს და ქმნის დარგთაშორისი კვლევის ინოვაციურ მეთოდებს.

1.1. ცნებისათვის „დიგიტალური ქართველოლოგია“

თუ თქვენ შეეცდებით სამეცნიერო ლიტერატურაში ან საინტერნეტო სივრცეში მოძებნოთ ამ ცნების – **დიგიტალური ქართველოლოგია** – განმარტება, ამაოდ გაირჯებით და არა მხოლოდ იმიტომ, რომ მეცნიერების ეს დარგი ჩვენ თვალწინ ყალიბდება, და შესაბამისად არ არსებობს მისი განსაზღვრებაც[57], არამედ იმიტომაც, რომ ამ შესიტყვებაში – **დიგიტალური ქართველოლოგია** – პრობლემას ქმნის თავად ცნება (და შესაბამისად ტერმინიც) „ქართველოლოგია“. დარგის ჩამოყალიბება და განვითარება ცნების განსაზღვრებასთან არის დაკავშირებული. ცნების დეფინიცია კი შეუძლებელია ოთხი არსებითი კომპონენტის – **დარგის არსის, კვლევის ობიექტის, მეთოდების განსაზღვრისა** და მეტაენის შექმნის გარეშე. ამთავითვე გვინდა აღვნიშნოთ, რომ აქ ორ არსებით პრობლემას ვაწყდებით, რომელიც უკავშირდება, ერთი მხრივ, ფენომენოლოგიურ პრობლემას (დარგის არსისა და საკვლევი ობიექტის საკითხს), მეორე მხრივ კი კლასიფიკაციის საკითხს.

---

[57] ტერმინი „დიგიტალური ქართველოლოგია“ პირველად ჩვენ მიერ იქნა გამოყენებული 2016 წელს ქართველი სრულდერებისათვის შექმნილი საავტორო ინტერნეტგვერდის „Digital Kartvelology“-ის ფარგლებში. ცნების პირველი დეფინიციაც ამ გვერდზე არის მოცემული. იხ.: http://tandaschwili.com/

საბჭოთა (მათ შორის ქართულ) სამეცნიერო სიერცესა და ევროპულ სამეცნიერო პრაქტიკაში **ქართველურ** და **კავკასიურ ენათა** განსხვავებული კლასიფიკაცია არის გავრცელებული. პირველი **ქართველურ ენებს** განიხილავს როგორც დამოუკიდებელ დისციპლინას, რომელიც კავკასიურ ენებთან (აფხაზურ-ადიღეურ, ნახურ და დაღესტნურ ენებთან) ერთად ქმნის **იბერიულ-კავკასიურ** ენათა ოჯახს, მეორე კი ქართველურ ენებს მთის იბერიულ-კავკასიურ (კავკასიაში გავრცელებულ ავტოქტონურ) ენებთან ერთად აერთიანებს **კავკასიურ ენათა ოჯახში** (Caucasian languages), რომელიც სამ ენობრივ ქვეჯგუფად არის დაყოფილი: აფხაზურ-ადიღეური ენები (West Caucasian languages), სამხრეთ-კავკასიური ანუ ქართველური ენები (South Caucasian / Kartvelian languages) და ნახურ-დაღესტნური ენები (North-East Caucasian/ Nakh-Daghestanian languages). შესაბამისად, ევროპული სამეცნიერო ტრადიციის მიხედვით კავკასიოლოგიაში იგულისხმება სამეცნიერო დისციპლინა, რომელიც ზემოთ აღნიშნულ ენათა ქვეჯგუფებს მოიცავს, მათ შორის, ქართველოლოგიასაც. ამ შემთხვევაში ქართველური ენები კავკასიური ენების პარადიგმაში მის ერთ-ერთ კომპონენტად არის მოაზრებული. ქართულ სამეცნიერო სიერცეში კი ქართველოლოგია კავკასიოლოგიის გვერდით ერთ სინტაგმატურ პარადიგმაში მოიხსენიება და დამოუკიდებელ დისციპლინას წარმოადგენს.

მეორე პრობლემა **ქართველოლოგიის** ფენომენოლოგიურ გაგებას უკავშირდება. აღნიშნული ტერმინი სამეცნიერო ლიტერატურაში ორ ცნებას მიემართება და, შესაბამისად, ორი – ვიწრო და ფართო მნიშვნელობით გამოიყენება. ტერმინი **„ქართველოლოგია“** ფართო მნიშვნელობით განიმარტება, როგორც **„ქართველთმცოდნეობა“** და მოიცავს ქართველური ეთნოსის, ენებისა და მათთან მიმართებაში არსებული კულტუროლოგიური ფენომენების კვლევას და აქ იგულისხმება ყველა ჰუმანიტარული დარგი – ენათმეცნიერება, ლიტერატურათმცოდნეობა, ეთნოგრაფია, ნუმიზმატიკა, ისტორია, არქეოლოგია, ეკონომიკა, სამართალმცოდნეობა, ეპიგრაფიკა, პალეოგრაფია, წყაროთმცოდნეობა და ა.შ. შესაბამისად, **ქართველოლოგია** ამ შემთხვევაში უნდა განვიხილოთ როგორც **ტრანსდისციპლინური ტერმინი** და მისი მოცულობა ფარავს ცნებას „ქართველთმცოდნეობა“[58]. თუმცა, სამეცნიერო მიმოქცევაში ხმარებული ტერმინი „ქართველოლოგია“ უმეტესად ვიწრო გაგებით გამოიყენე-

---

[58] იხ. ტერმინის განმარტება უნივერსალურ ენციკლოპედიურ ონლაინლექსიკონში – *ქართველთმცოდნეობა - ქართული ენისა და კულტურის (ლიტერატურის, ისტორიის, ეთნოგრაფიის) შემსწავლელი მეცნიერებათა ერთობლიობა.*
http://www.nplg.gov.ge/gwdict/index.php?a=term&d=14&t=198327

ბა და აღნიშნავს **ქართველური ენების – ქართულის, ზანურისა და სვანურის –** შემსწავლელ მეცნიერებას. ასეთ შემთხვევაში **ქართველოლოგია** უნდა განვიხილოთ, როგორც ლინგვისტური ტერმინი. იმისდა მიხედვით, თუ „ქართველოლოგიის" რომელ განსაზღვრებას ვიზიარებთ, განსხვავებულად განისაზღვრება მოცემული დისციპლინის როგორც არსი, ისე კვლევის ობიექტი, მეთოდოლოგიური და თეორიული ჩარჩო. ამ ტერმინოლოგიური ომონიმიის მოსახსნელად აუცილებელია ცალკე ტერმინი შეიქმნას **ქართველოლოგიის, როგორც ტრანსდისციპლინური ტერმინის** გამოყენების შემთხვევაში, რათა განვასხვაოთ იგი **ქართველოლოგიის, როგორც ლინგვისტური ტერმინისაგან,** როგორც ეს, მაგალითად, ინგლისურ ენაში არის დიფერენცირებული: **Kartvelology** *vs.* **Kartvelian studies.** თუ რამდენად მისაღებია ჩვენ მიერ შემოთავაზებული ტერმინი **ქართველთმცოდნეობა,** ამას უახლესი მომავალი გვიჩვენებს[59].

1.2. ქართული ენის კომპიუტერიზაცია

დიგიტალური ქართველოლოგია 21-ე საუკუნის პირმშოა და მისი ჩასახვა, ფორმირება და განვითარება ორ გეოგრაფიულ გარემოში – საკუთრივ საქართველოში და ევროპაში, კერძოდ გერმანიაში – განხორციელდა და ორი სხვადასხვა მიმართულებით განვითარდა: ა) ენის მოდელირების პრობლემების გადაჭრისა და ბ) დიდ მონაცემთა ბაზების შექმნის გზით.

დიგიტალური ქართველოლოგიის ჩანასახები, დიგიტალური ჰუმანიტარიის მსგავსად, უნდა განვიხილოთ როგორც **ქართული ენის კომპიუტერიზაციის** პროცესი, როდესაც საქართველოში ჯერ კიდევ გასული საუკუნის 60-იან წლებში საფუძველი ეყრება გამოყენებითი ლინგვისტიკის სფეროში ქართული ენის ფორმალურ დამუშავებას. ენის ავტომატური ანალიზისა და სინთეზის აქტუალობა იმ პერიოდისათვის ინოვაციური თეორიული ნაშრომების შექმნით – ნოამ ჩომსკისა (Syntactic Stuctures) და რიჩარდ მონტეგიუს ნაშრომებით (English as a Formal Languege, The Proper Treatment of Quantification in Ordinary Inglish) იყო განპირობებული. სწორედ ამ ნაშრომებმა მისცეს ბიძგი ახალი – ლოგიკურ-ლინგვისტური სამეცნიერო მიმართულების „ენა და ლოგიკა" ჩამოყალიბებას და განვითარებას. ეს სამეცნიერო სფერო ბოლო ორმოცი წლის განმავლობაში განსაკუთრებით სწრაფად განვითარდა ისეთი ფუნდამენტური თეორიული ნაშრომების შედეგად, როგორიცაა „ენა და ლოგიკა" ("Logic and Language", Edited by: Johan van Benthem and Alice ter Meulen, 1997), „თანამედ-

[59] ინტერნეტსივრცეში ჯერჯერობით მხოლოდ 260-ჯერ დასტურდება აღნიშნული ტერმინი. სტატისტიკური მონაცემის თარიღი: 3.01.2021.

როვე სემანტიკური თეორიები" ("Contemporary Semantic Theory", Handbook of Contemporary Semantic Theory (Blackwell Handbooks in Linguistics, 1997) და „ფორმალური სემანტიკები" ("Formal Semantics", Paul Portner, Barbara Partee, 2001)[60].

მე-20 საუკუნის ბოლოს ქართული სამეცნიერო საზოგადოება (როგორც ლინგვისტები, ისე ინფორმაციული ტექნოლოგიების წარმომადგენლები), ეპოქალური გამოწვევების წინაშე დადგა, რომელიც რამდენიმე ინტერდისციპლინური ხასიათის ამოცანის გადაწყვეტას მოითხოვდა. ეს ამოცანები რამდენიმე პუნქტად შეგვიძლია გამოვაყალიბოთ:

ა) ქართული ენის კომპიუტერიზაცია აუცილებლობით მოითხოვდა ნორმირებული ქართული ენის ლოგიკურ-მათემატიკურ ფორმალიზაციას მისი სამეცნიერო და გამოყენებით ტექნოლოგიებში ჩანერგვის მიზნით;

ბ) სამომხმარებლო სფეროში ენის ტექნოლოგიების გამოყენების აუცილებლობამ წინა პლანზე წამოსწია სპიკერ-მეტყველების (speech) მანქანური დამუშავების მნიშვნელობა – საჭირო გახდა ტექსტის მეტყველებად გარდაქმნის და პირუკუ პროცესის ტექნოლოგიზება (text to speech *vs.* speech to text technology). ქართული ენის სრულფასოვანი ფუნქციონირებისათვის „ტექნოლოგიების ბუმის ეპოქაში" აუცილებელ და გადაუდებელ ამოცანას წარმოადგენდა მეტყველების ამოცნობის ფორმალური სისტემის დამუშავება.

თეორიულ-კვლევითი სამუშაოები ამ მიმართულებით ძირითადად მეცნიერებათა აკადემიის ფარგლებში – არნ. ჩიქობავას სახელობის ენათმეცნიერების ინსტიტუტსა და მართვის სისტემების ინსტიტუტში – მიმდინარეობდა. ამ პერიოდის მეცნიერთაგან უნდა დაეასახელოთ ზურაბ ჯაფარიძე, თედო უთურგაიძე, ლევან ჩხაიძე, გიორგი ჩიკოიძე, ჯემალ ანთიძე, რომლებიც მცირე სამეცნიერო ჯგუფებთან ერთად ცდილობდნენ აღნიშნული პრობლემების გადაჭრას.

დამოუკიდებელი საქართველოს დაარსების შემდეგ, 90-იანი წლების ბოლოს შესაძლებელი გახდა საბჭოთა პერიოდისათვის დამახასიათებელი „ჩაკეტილობის" გადალახვა და უცხოელ მეცნიერებთან თანამშრომლობა. გახშირდა საერთაშორისო კონფერენციები ენისა და ლოგიკის, ენის ავტომატური მოდელირების, ენის, ლოგი-

---

[60] იხ. კ. ფხაკაძე. წინათქმა ანუ ლოგიკისა და ენის გაერთიანებული ქართული ჯგუფის მისწრაფებისა და ამოცანების შესახებ, სამეცნიერო-საგანმანათლებლო ჟურნალი „ქართული ენა და ლოგიკა", თბილისი, 2005, იანვარი-ივნისი, გვ. 5-7.
https://gcoanbani.com/other/pdfs/Publications/magazine1.pdf

კისა და კომპიუტერიზაციის საკითხებზე[61].

მიუხედავად იმისა, რომ ქართული ენის კომპიუტერიზაციის მიმართულებით 60-იან წლებში დაიწყო მუშაობა, ისეთი სრულფასოვანი შედეგი, როგორიც, მაგალითად, ევროპული ენებისათვის იყო მიღწეული, ქართული ენის შემთხვევაში სამწუხაროდ ვერ განხორციელდა. ამის მიზეზად, თავად ამ დარგში მომუშავე ავტორების აზრით, დასახელებულია ის ფაქტი, რომ ამ საკითხზე მუშაობის დროს მეცნიერები ყურადღებას ძირითადად ამახვილებდნენ პრობლემის შიდაენობრივ ფორმალურ გადაწყვეტაზე და იგნორირებული იყო ლოგიკურ-ლინგვისტური მიდგომა, კერძოდ პრობლემის სემანტიკური ასპექტი, რომელიც შინაარსის ფორმალიზაციისათვის (ერთი ენიდან მეორე ენაზე გადატანის მიზნით), მათემატიკურ ენას იყენებს, როგორც შინაარსის ერთი ენიდან მეორეზე გადატანის უნივერსალურ ენობრივ საშუალებას[62]. თუმცა აქვე უნდა დავასახელოთ ის ინსტიტუციები და მიღწევები, რომელიც დღეისათვის არსებობს ქართული ენის ტექნოლოგიზაციის თვალსაზრისით[63]. ესენია:

- საქართველოს ტექნიკური უნივერსიტეტის არჩილ ელიაშვილის სახელობის მართვის სისტემების ინსტიტუტის ენობრივი მოდელირების განყოფილება (ხელმ.: გიორგი ჩიკოიძე 2018 წლამდე, 2018 წლიდან - ლიანა ლორთქიფანიძე), რომელმაც თეორიული ნაშრომების გარდა[64] პრაქტიკულად მოქმედი პროგრამების განხორციელება შეძლო, კერძოდ: ქართულ სიტყვათა ქსელის კომპაილერი (1915-2016, ხელმძ.: ლიანა ლორთქიფანიძე); ქართული ენის კორპუსის სრული (მორფოლოგიური, სინტაქსური, სემანტიკური) ანოტირების სისტემა (2013-2016, ხელმძ. გიორგი ჩიკოიძე); ქართული ენის კომპიუტერული სუფლიორი უნარდაქვეითებულ პირთათვის (2009-2011, ხელმძ.: ლიანა ლორთქიფანიძე,

---

[61] არნ. ჩიქობავას სახელობის ენათმეცნიერების ინსტიტუტში 1996-1998 წლებში ტარდებოდა სამეცნიერო კონფერენცია თემაზე: „ენის კონცეპტუალური და კომპიუტერული მოდელები" (ორგანიზატორები: მანანა თანდაშვილი და ირაკლი ქოწენიძე), 1995 წლიდან საქართველოში სისტემატურად იმართება საერთაშორისო სიმპოზიუმი International Tbilisi Symposium Series on Language, Logic and Computation, რომელსაც ამსტერდამის უნივერსიტეტი ატარებს თსუ-სა და საქართველოს მეცნიერებათა აკადემიასთან ერთად. იხ. http://events.illc.uva.nl/Tbilisi/

[62] იხ. სამეცნიერო-საგანმანათლებლო ჟურნალი „ქართული ენა და ლოგიკა", ლოგიკისა და ენის გაერთიანებული ქართული ჯგუფი, 2005
https://geoanbani.com/other/pdfs/Publications/magazinc1.pdf

[63] საგანგებოდ გვსინდა აღვნიშნოთ, რომ ჩვენს მიზანს წინამდებარე ნაშრომში არ წარმოადგენდა ქართული ენის ტექნოლოგიების საკითხებზე მომუშავე ჯგუფებისა და პროდუქტების აღრიცხვა-დახასიათება; შესაბამისად, ჩამონათვალი არასრულია.

[64] გ. ჩიკოიძე, ლ. მარგველანი, ლ. თამარაშვილი, ბუნებრივი ენა და გამოთვლითი მანქანა, თბილისი, მეცნიერება, 1987.

პროგრამა დაინერგა თბილისის 203-ე ყრუ და სმენადაქვეითებულთა საჯარო სკოლა-პანსიონში); ავტომატური განმარტებით-კომბინატორული ლექსიკონი როგორც ქართული ენის მოდელირების საფუძველი (2009-2011, ხელმძ. გიორგი ჩიკოიძე); გიორგი ჩიკოიძე წლების განმავლობაში თანამშრომლობდა მონრეალის უნივერსიტეტის ლინგვისტიკისა და თარგმნის დეპარტამენტის პროფესორ იგორ მელჩუკთან და მოსკოვის ლინგვისტური სკოლის წარმომადგენლებთან იური აპრესიანთან, ნინა ლეონტიევასთან, იგორ ბოგუსლავსკისთან. 2005-2007 წლებში შვედეთის სამეფო აკადემიის პროფესორ დოქტორ შერი ჰანიკატთან ერთად დაამუშავა პროექტი „რუსული ტექსტის ბეჭდვის კომპიუტერული მხარდაჭერა“.

- ქართული ენობრივი ტექსტების მორფოლოგიური სპელჩეკერი, რომელიც წარმოქმნის პრინციპზე აგებულ ქართული მორფოლოგიური მართლმწერის ელექტრონულ პროგრამულ სისტემას წარმოადგენს და პირველად რეალიზებულ იქნა 2003 წელს. პროგრამა დამუშავდა ლოგიკისა და ენის გაერთიანებული ქართული ჯგუფის მიერ (ხელმძღვანელი კ. გაბუნია). ელექტრონული პროგრამული სისტემა არა ამომწურავი სისრულით, მაგრამ, საკმარისი წარმადობით ართმევს თავს ქართული ენობრივი ტექსტების მორფოლოგიური გასწორების ამოცანას.
- 1998 წელს დაფუძნებული „ქართული ინფორმატიკულ-ლინგვისტური ცენტრი“ (ქ. დათუკიშვილი, ნ. ლოლაძე, მ. ზაკალაშვილი), რომელიც 2009 წლიდან ფუნქციონირებს „ლინგვისტური ტექნოლოგიების ჯგუფის“ სახელით. ლინგვისტებისა და პროგრამისტებისაგან შემდგარი გუნდი მუშაობს ენათმეცნიერებისა და თანამედროვე ტექნოლოგიების მიღწევების გამოყენებით ქართული ენის ტექნოლოგიების შექმნაზე. დღეისათვის ჯგუფის მიერ შექმნილია „ქართული ლექსიკონი“ (ბეჭდური გამოცემის ელექტრონული ვერსია ეყრდნობა ტექსტების კორპუსს და მოიცავს 41302 სიტყვა-სტატიას) და ენობრივ შეცდომათა ელექტრონული ლექსიკონი, რომელიც მასმედიის, ტელეკომპანიებისა და აბიტურიენტთა ნაშრომების კომპიუტერული ანალიზის შედეგად შეკრებილ ტიპობრივ შეცდომათა ელექტრონულ ბაზას წარმოადგენს. ასევე, ტექნოლოგიური პროდუქტები: ქართული ენის მორფოლოგიური პროცესორი და პროგრამა "ლექსიკოგრაფი".
- ქართული ენის სინტაქსის მასწორებელი სისტემის პირველი ექსპერიმენტული ვარიანტი, რომელიც ლოგიკისა და ენის ქართულ ჯგუფში კ. ფხაკაძის ხელ-

მძღვანელობით შემუშავდა. კ. ფხაკაძის თეორიული ხედვების საფუძველზე შემუშავებული ალგორითმი ტიპიზირებული მათემატიკური ენის მეშვეობით ამოიცნობს შანიძისეული მარტივი წინადადებებისა და სახელური ფრაზებისათვის დამახასიათებელ არასწორ სინტაქსურ კონსტრუქციებს და ასწორებს მათ. პროგრამული მოდული აგებულია მხოლოდ მათემატიკური მიდგომებით და პროგრამული რეალიზაციისას არ იყენებს არცერთ გრამატიკულ წესს. გრამატიკული წესების ნაცვლად სისტემა ეყრდნობა ტიპიზირებული მათემატიკური ენის ფარგლებში მათემატიკური ფორმულების სახით ჩაწერილ გამოსახულებათა გადამწერ წესებს (პროგრამული გადაწყვეტა ეკუთვნის რ. სხირტლაძეს).

- ქართული ენის ზეპირმეტყველების ავტომატური ამოცნობის ექსპერიმენტული პროგრამა, რომელიც გ. ჩიჩუას მიერ წარმოებულ ინტენსიურ კვლევებს ემყარება და რომელსაც მეტყველების ამოცნობის პრობლემაზე ორიგინალური თეორიული მიდგომებთან ერთად უკვე აქვს ამოცნობის მაღალმაჩვენებლიანი ექსპერიმენტული პროგრამული რეალიზაციები.
- ქართული მოდელირების ენის ასოციაცია, რომლის მიზანია ქართული ენის მოდელირება, ენობრივი სისტემის ფუნდამენტური კვლევა და კომპიუტერული დამუშავება. ასოციაციამ გააციფრა ქართული ენის ორთოგრაფიული და განმარტებითი ლექსიკონები, რომლებიც ღია რესურსის სახით არის განთავსებული ინტერნეტში[65].
- თბილისის სახელმწიფო უნივერსიტეტის დასავლეთ ევროპის ენებისა და ლიტერატურის ფაკულტეტის ლექსიკოგრაფიული ცენტრი პროფ. თინათინ მარგალიტაძის ხელმძღვანელობით წლების განმავლობაში ამუშავებს და ხვეწს ბილინგვური და მულტილინგვური დიგიტალური ლექსიკონის გაწყობის თეორიულ-სტრუქტურულ პრინციპებსა და ტექნოლოგიურ ჩარჩოს. ცენტრის მიერ შექმნილმა ტექნოლოგიურმა პროდუქტმა – დიდი ინგლისურ-ქართული ლექსიკონი (CEGD), რომელიც 110 000 სალექსიკონო ერთეულის რამდენიმე ასეულ ათას ინგლისურ მნიშვნელობას, კოლოკაციებს, ფრაზულ ზმნებს, იდიომებსა და ტერმინებს მოიცავს, საქართველოში დიგიტალური ლექსიკოგრაფიის განვითარებას ჩაუყარა საფუძველი

ევროპაში ქართული ენის კომპიუტერიზაცია უკავშირდება უპირველეს ყოვლისა გერმანელი ქართველოლოგის და კავკასიოლოგის პროფ. იოსტ გიპერტის სა-

[65] იხ. http://cna.ge/

ხელს, რომელმაც გასული საუკუნის 80-იან წლებში დაიწყო ქართული ენის კომპიუტერიზაციის მიმართულებით მუშაობა, ოღონდ ქართველი კოლეგებისაგან განსხვავებით, ეს პროცესი სრულიად სხვა მიმართულებით წარმართა, კერძოდ, **ქართველოლოგიის დიგიტალიზაციის, კერძოდ კი დიდ მონაცემთა ბაზების შექმნის გზით.** დიგიტალური ქართველოლოგიის განვითარებაში გერმანელი ქართველოლოგის წელილის შესახებ საგანგებოდ ვისაუბრებთ ქვემოთ.

1.3. ქართველოლოგიის დიგიტალიზაციის ეტაპები

ენის კომპიუტერიზაციის პროცესი, ზოგადად, გლობალურ სამეცნიერო სივრცეში, ორი მიმართულებით წარიმართა: **ენობრივ მონაცემთა დაგროვებისა და ენის მოდელირების** გზით. ენობრივ მონაცემთა დაგროვება ეყრდნობოდა **დაგროვების პრინციპს**, რაც მოცემული ენისათვის **დიდ მონაცემთა ბაზების** შექმნას გულისხმობს და მყარ ემპირიულ საფუძველს უქმნის დარგის თეორიულ განვითარებას. ეს პრინციპი ეფექტიანია ისეთი ენებისათვის, რომელთა მორფოლოგია შედარებით მარტივი ფლექსიით ხასიათდება. მაგალითად, დაგროვების პრინციპი წარმატებით იქნა გამოყენებული ინგლისური ენისთვის, რომელსაც არ გააჩნია ბრუნების სისტემა, ხოლო უღელილების სისტემა ძირითადად აღწერითი ფორმების წარმოებას ემყარება, რაც კვანტიტატიური თვალსაზრისით შედარებითი საზღერულობით ხასიათდება. რთული მორფოლოგიური სისტემების მქონე ენებისათვის, როგორიცაა, მაგალითად, ქართული, ქართველი მეცნიერების მიერ თავიდანვე იქნა უარყოფილი „დაგროვების" პრინციპის მეთოდის გამოყენება. პროფ. თედო უთურგაიძის თეორიული ნაშრომების მიხედვით ქართული ენის გარდამავალი სამპირიანი ზმნის პარადიგმა ერთი მწკრივის ფარგლებში 112 ორგანული წარმოების ფორმას გვიჩვენებს, ორპირიან ზმნებში ერთი სერის ფარგლებში - 335 ფორმას, ხოლო ყველა სერში - 2010-ს. ამას ემატება ე.წ. მორფოლოგიური ომონიმია (მაგ.: *გაქებთ* ფორმის მორფოსინტაქსური ანალიზი თ. უთურგაიძის მოდელის მიხედვით ოთხ სცენარს შეესაბამება: *გაქებთ მე თქვენ. გაქებთ ის თქვენ. გაქებთ ჩვენ შენ. გაქებთ ჩვენ თქვენ*), რაც მნიშვნელოვნად ართულებს ფორმალიზაციის ამსახველი ალგორითმის შექმნას. კიდევ უფრო გართულდება სურათი, თუ ზმნის უღელილების ყველა მწკრივის ფორმას ავიღებთ. ასეთ შემთხვევაში გარდამავალი ზმნის ფორმათა რაოდენობა უღელილების პარადიგმაში 2000-ს გადააჭარბებს[66]. აღნიშნული კვანტიტატიური ხასიათის სირთულეები, რომელიც ქართუ-

[66] თედო უთურგაიძე. „გრამატიკული კატეგორიებისა და მათი ურთიერთმიმართებისათვის ქართულ ზმნაში", თბილისი, 2002, გვ. 112, 146.

ლი ენისათვის დიდი მონაცემების შექმნისა და გამოყენების არაეფექტიანობას „წინასწარმეტყველებდა“. გახდა იმის მიზეზი, რომ საქართველოში ქართული ენის კომპიუტერიზაცია ენობრივი სისტემის ფორმალიზაციის – ენის მოდელირების მიმართულებით წარიმართა. 90-იან წლებში დამუშავდა არაერთი თეორია და შეიქმნა არაერთი სამეცნიერო ჯგუფი, რომელიც ენის მოდელირების გზით ცდილობდა პრობლემის გადაჭრას, თუმცა ენის ავტომატური სინთეზისა (და შესაბამისად, ანალიზის) ტექნოლოგიზება ვერ განხორციელდა; ან განხორციელდა, მაგრამ ვერ მოხერხდა შექმნილი ტექნოლოგიების ვერიფიცირება – პროგრამების გამოცდა **დიდ მონაცემთა ბაზებზე.**

დიდ მონაცემთა ბაზების შექმნა საუკუნეთა მიჯნაზე განხორციელებული საყოველთაო ტექნოლოგიზაციის ბუმით იყო განპირობებული. ენობრივი რესურსების გაციფრება, ანუ ენობრივი ინფორმაციის ახალ მატარებლებზე გადატანა, ზუსტად შეესაბამებოდა ეპოქის მოთხოვნებს. კომპიუტერის გლობალურმა გამოყენებამ უალტერნატივო გახადა ინფორმაციის გაციფრების პროცესი და სრულიად განდევნა „საბეჭდი მანქანა“, რამაც, შესაბამისად, ენათმეცნიერები კვლევის თანამედროვე, ციფრულ მატარებლებზე ორიენტირებული საშუალებების განვითარებისა და ელექტრონული რესურსების შექმნის აუცილებლობის წინაშე დააყენა. ევროპულ ქართველოლოგიაში ამ ორი – „დაგროვების პრინციპისა“ და „ენის მოდელირების პრინციპის“ სინთეზის შედეგად საფუძველი ჩაეყარა დიგიტალური ქართველოლოგიის განვითარებას და გერმანელი ქართველოლოგის, იოსტ გიპერტის სახელთან არის დაკავშირებული.

იოსტ გიპერტი 80-იან წლებში დაინტერესდა ენის ტექნოლოგიებით და ჯერ კიდევ ბერლინის თავისუფალ უნივერსიტეტში იწყებს ენათა კვლევის თანამედროვე მეთოდების კონცეპტუალიზაციას. იოსტ გიპერტის კონცეპტუალური მიდგომა ორ არსებით ფაქტორს ემყარებოდა – **დიდ მონაცემთა ბაზების შექმნას** და **ენის ავტომატური დამუშავების ტექნოლოგიების განვითარებას,** რომელიც წარმატებით განახორციელა საერთაშორისო პროექტის TITUS-ის (Thesaurus Indogermanischer Text- und Sprachmaterialien) პლატფორმის ფარგლებში.

იოსტ გიპერტის მოღვაწეობა ქართული და კავკასიური ენის დიგიტალიზაციის მიმართულებით რამდენიმე ეტაპად შეიძლება დაყოთ. პირველ ელექტრონულ რესურსებს ქართული ენისათვის იოსტ გიპერტი ინდოევროპული ენების ელექტრონული პლატფორმის TITUS-ის ფარგლებში ქმნის. ამ პერიოდში იგი აქტიურად თანამშრომლობს ივ. ჯავახიშვილის სახ. თბილისის სახელმწიფო უნივერსიტეტსა და

სხვა უმაღლეს სასწავლებლებში მოღვაწე ქართველ მეცნიერებთან. განსაკუთრებით მჭიდროდ სამეცნიერო ურთიერთობა აქვს მას პროფ. ზურაბ სარჯველაძესთან და პროფ. ვახტანგ იმნაიშვილთან და მათთან ერთად განავრცობს მის მიერ შემქნილ და TITUS-ის პლატფორმაზე განთავსებულ ძველი და ახალი აღთქმის წიგნებისა და ჰაგიოგრაფიული მწერლობის ნიმუშების კოლექციას.

დიგიტალიზაციის ამ ეტაპზე იოსებ გიპერტი ნაწილობრივ ენობრივი რესურსების მზა ტექნოლოგიებს იყენებდა, ნაწილობრივ კი თავად ქმნიდა მათ.

```
!bŠušaniķ!P11!!1  [735] ttuesa Oķdonbersa: iz~;
!12  çamebay çmidisa Šušaniķisi dedoplisay»
!13  da aç damţķicebulad gitxra tkuen aġsasruli çmidisa da sanaţrelisa
!14  Šušaniķisi.
!15  1. iǫo mervesa çelsa [Mr...] sṗarsta mepisasa ķarad samepod
!16  çaremarta Varsken ṗlţiaxši, zē Aršušaysi, rametu ṗirvel igi-ca
!17  iǫo kristţlane, našobi mamisa da dedisa kristţianetay. da colad
!18  misa iǫo asuli Vardanisi, somexta sṗaypēţisay, romlisatws ese mivçere
!19  tkuenda, mamisagan saxelit Vardan, da siǫuarulit saxeli
!110  misi Šušaniķ, mošiši ġmrtisay, vitarca-igi vtkut, siǫrmitgan twsit.
!111  da uķeturebisa mis saxisa kmrisa misisaysa maradis guls eţǫwn da
!112  evedrebin ǫovelta-ve locvis-ǫopad mistws, rayta-mca cvala igi
!113  ġmertman saxisa misgan ugunurebisa da ikmna-mca gonierebasa
!114  Kristţēssa.
!115  xolo amis ubadruķisa da sam gzis saçǫalobelisa Varskenisi
!116  vitar-me vin tkuas ǫovlad gançirulisay, vitar-igi ganagdo sanaţreli
!117  sasoebay Kristţēsi?! anu vin ara sţirodis mas, romelsa
!118  arca çiri exliva, arca šiši, arca maxwli, arca ṗǫrobilebay
!119  Kristţēstws?!
!120  rametu ražams çardga igi çinaše sṗarsta mepisa, ara tu ṗaţivisa
!121  moǫebisatws, aramed zǧunad tavsa twssa šesçirvida mepisa
!P12!!1  mis mimart uvaris-ǫopita češmariţisa ġmrtisayta da taǫuanis-scemda
!12  cecxisa [736], rametu ǫovlit ķerzo-ve tavi twsi Kristţēsgan ganagdo.
DOS G:\BYB\AGEO\GH\GH.BYB                    Dat 2   S 17   Z  12      Pos 3
```

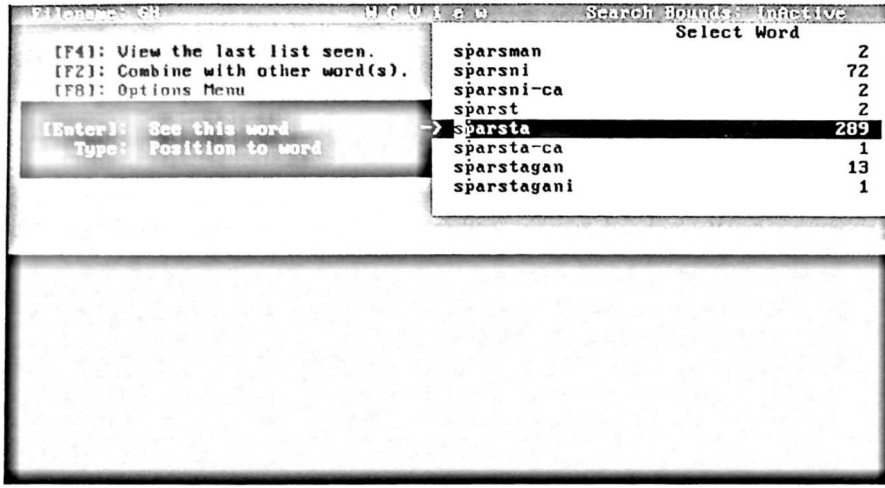
„შუშანიკის წამების“ სტრუქტურირებული ტექსტი DOS-ის ფორმატში

დიგიტალიზაციის შემდგომ ეტაპზე გიპერტი იყენებს ენობრივ მონაცემთა საგანგებო ოპერაციულ სისტემას WordCruncher-ს, რომელშიც ტექსტები ინტეგრირებული იყო სტრუქტურირებული სახით, რაც ძიებისა და კონკორდანსის გენერირების საშუალებას იძლეოდა:

```
                    Šušaniķ  Heading                Expanded Window
 26  xar cata šina. "ġmerti Israēlisay, man mosces ǯali da simţķice
 27  ersa twssa"». Baruk aqsena Israēli da simţķice, rametu vitarca sasovebay
 28  mamuli da galobay. da simţķicē amaoy ganišora da Israēli mţķiced
 29  šeiçqnara, esret saxed-ve ikmna šoris ķrebulsa ǯuelisa da axalisa
 30  erisasa. ķurtxeul ars ġmerti? raoden šuenierad aġesrula dasasruli
 31  psalmunisay, rametu iţǫws: romelman exeodenni ķetilis-ǫopani
 32  mohmadlna natesavsa ķactasa, romel miutxrobeli da miuçdomeli twsi
 33  ķact-moǫuarebay ačuena da ǫo macxovarebay ǫovelta ķideta, martal da
 34  ǧirs ars kebad da didebad da ķurtxevad ġmerti maǧali. amen.
 35

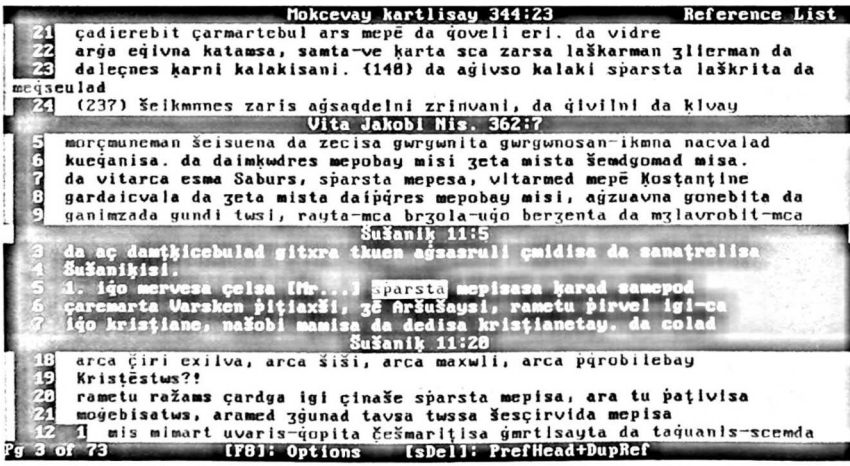
Šušaniķ  11  1  [735] ttuesa Oķdonbersa: iz~;
  2  çamebay çmidisa Šušaniķisi dedoplisay»
  3  da aç damţķicebulad gitxra tkuen aġsasruli çmidisa da sanaţrelisa
  4  Šušaniķisi.
  5  1. iǫo mervesa çelsa [Mr...] sṗarsta mepisasa ķarad samepod
  6  çaremarta Varsken ṗlţiaxši, zē Aršušaysi, rametu ṗirvel igi-ca
  7  iǫo kristţlane, našobi mamisa da dedisa kristţianetay. da colad
  8  misa iǫo asuli Vardanisi, somexta sṗaypēţisay, romlisatws ese mivçere
  9  tkuenda, mamisagan saxelit Vardan, da siǫuarulit saxeli
 10  misi Šušaniķ, mošiši ġmrtisay, vitarca-igi vtkut, siǫrmitgan twsit.
 11  da uķeturebisa mis saxisa kmrisa misisaysa maradis guls eţǫwn da
 12  evedrebin ǫovelta-ve locvis-ǫopad mistws, rayta-mca cvala igi
[1] Mark Block
```

„შუშანიკის წამების“ სტრუქტურირებული ტექსტი WordCruncher-ში (1993)



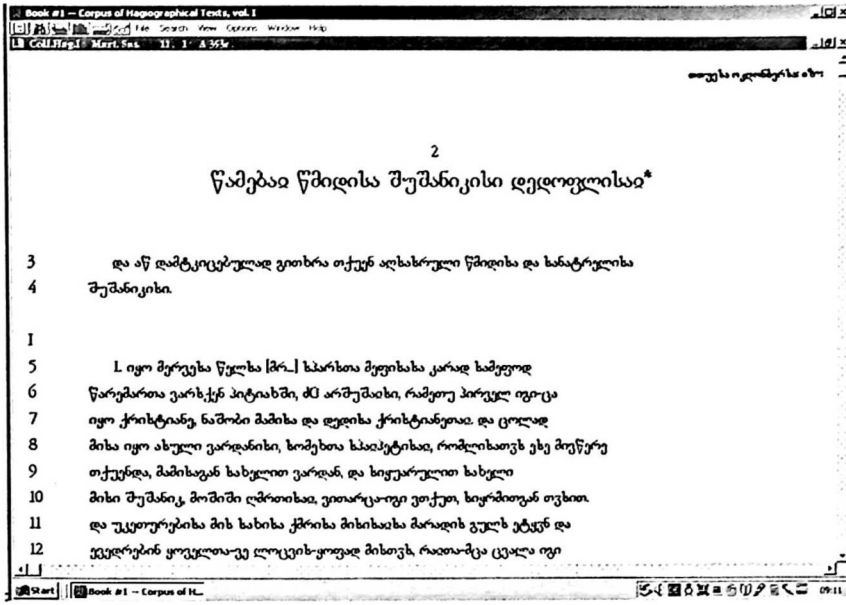
საძიებო სისტემა: ძიებაშია სიტყვაფორმა *სპარსთა*

კორპუსი აღჭურვილი იყო დიგიტალური წესების ბაზით, რომლის მეშვეობითაც კორპუსმენეჯერი კონკრეტულ ტექსტში არა მხოლოდ ცალკეულ სიტყვაფორმას, არამედ საძიებო სიტყვის ყველა შესაძლო – ფლექსიურ და დერივაციულ ფორმას ეძებდა და პოულობდა. იხ.: სხვადასხვა ტექსტებში სიტყვაფორმის „*სპარსი*“ ძიების შედეგად წარმოქმნილი სიტყვაფორმების სია (ზედა სურათი) და მოძიებული ფორმების ჩამონათვალში სიტყვაფორმის „სპარსთა“ მონიშვნის შედეგად მოძიებული კონტექსტები კონკორდანსის ფორმატში (ქვედა სურათი):



ძიების შედეგები, მოცემული კონკორდანსის ფორმით

შდრ.: WordCruncher-ის მოგვიანო ვერსია (1995-1999 წლები):



2

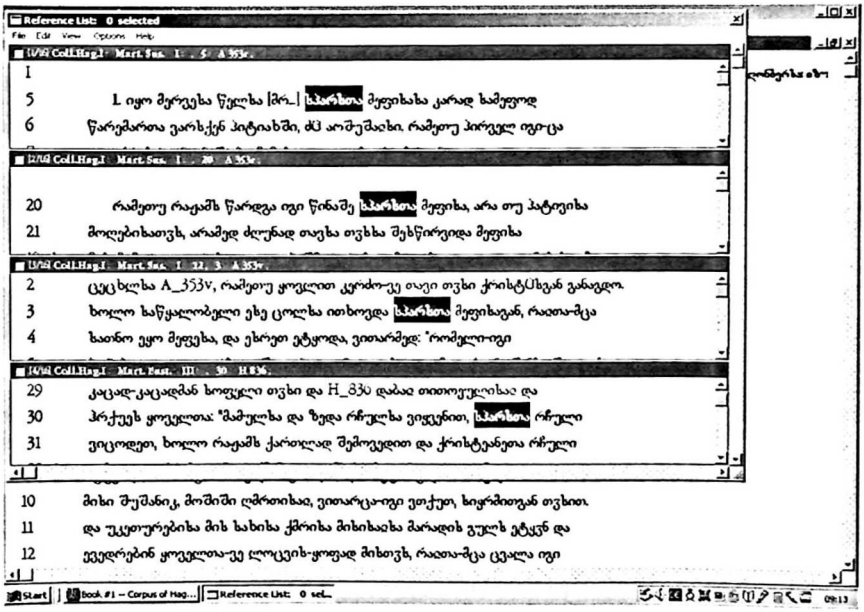
წამებაჲ წმიდისა შუშანიკისი დედოფლისაჲ*

და აწ დამტკიცებულად გითხრა თქუენ აღსახრული წმიდისა და სანატრელისა
შუშანიკისი.

I

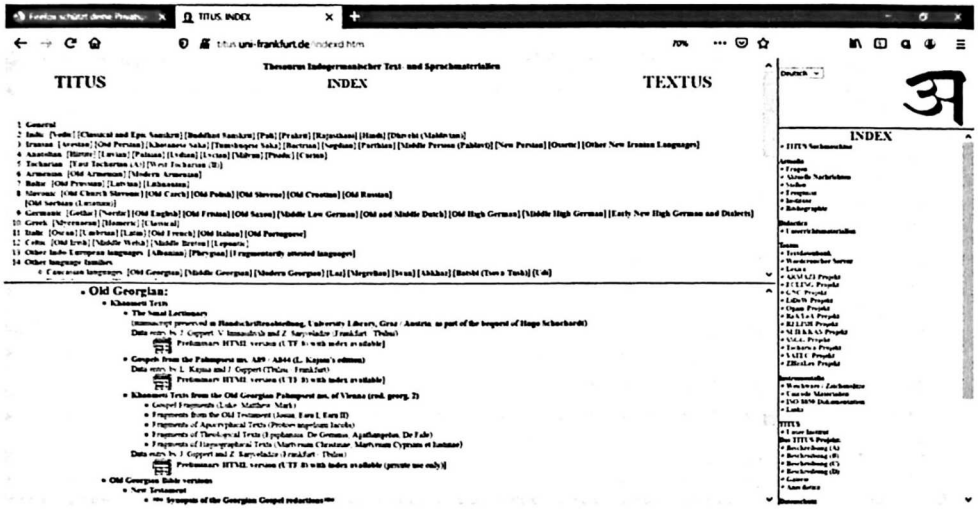
L იყო მერეჳა წელსა |მრ.| ხაზრთა მეფისახა კარად სამეფოდ
წარემართა ვარსქენ პიტიახში, ძჲ არშუშაისი, რამეთუ პირველ იგი-ცა
იყო ქრისტიანე, ნაშობი მამისა და დედისა ქრისტიანეთაჲ. და ცოლად
მისა იყო ახული ვარდანისი, ხოშეხთა ხაზაპეტისაჲ, რომელისათჳს ესე მუნერე
თქუენდა, მამისაგან სახელით ვარდან, და ხიყოარულით სახელი
მისი შუშანიკ, მოშიში ღმრთისაჲ, ვითარცა-იგი ეთქუე, ხიყრმითგან თჳსით.
და უკეთურებისა მის სახისა ქმრისა მისისაჲსა მარადის გულს ეტკჳნ და
ეგედრებინ ყოველთა-ვე ლოცვის-ყოფად მისთჳს, რაჲთა-მცა ცვალა იგი

„შუშანიკის წამების“ სტრუქტურირებული ტექსტი WordCruncher-ში



ძიების შედეგები, მოცემული კონკორდანსის ფორმით

დიდ მონაცემთა ბაზების შექმნის შემდგომ ეტაპზე განხორციელდა ქართველური ენების (და არა მარტო ქართველური ენების) ინტეგრაცია ენობრივი რესურსების დამუშავების ფართომასშტაბიან პორტალზე TITUS:



http://titus.uni-frankfurt.de/indexd.htm

ქართული ენის დიგიტალიზაციის პარალელურად იოსტ გიპერტი გასული საუკუნის 80-იან წლებში მეუღლესთან, სონია ფრიცთან ერთად აგროვებდა მასალებს კავკასიური ენების ელექტრონული რესურსების შესაქმნელად. მართალია, კავკასიურ ენათა დიგიტალიზაცია ამ პერიოდში არასისტემური იყო და ფრაგმენტულ ხასიათს ატარებდა, მაგრამ იგი რესურსების მრავალფეროვნებით გამოირჩეოდა, უპირატესად კი იმის გამო, რომ ტექსტური ფორმატის გარდა, აუდიოფორმატსაც მოიცავდა შემდეგი ენებისათვის: ქართული, მეგრული, ლაზური, სვანური, აფხაზური, ადიღეური, ხუნძური, ახვახური, კარატული, ბაცბური, თაბასარანული, აღულური, უდიური, დარგუული. ეს ჩანაწერები ძირითადად იოსტ გიპერტისა და სონია ფრიცის მიერ არის მოპოვებული 1984 წელს მათი საქართველოში ყოფნის დროს, თუმცა ასევე მოიცავს გასული საუკუნის დასაწყისში ჩაწერილ მასალებსაც, მათ შორის, ადოლფ დირის მიერ 1901 ჩაწერილ აუდიომასალას უდიური ენისათვის (მოქმელი ვ. გუკასიანი), რომელიც იოსტ გიპერტმა ვენის არქივში აღმოაჩინა, გააციფრა და ტიტუსის პორტალზე განათავსა[67].

[67] 1984 წელს იოსტ გიპერტის მიერ თბილისში ჩაწერილი რამდენიმე კავკასიური ენის აუდიო მასალები (მოქმელები: ზაჰიდათ მაჰომედბეკოვა, ალექსანდრე მაჰომედოვი, ელენე ქადაგიძე, ელისაბედ გაზდელიანი, წათე ბაწაში და სხვ.), ისევე როგორც ვენის არქივში მოპოვებული



იოსტ გიპერტისა და სონია ფრიცის მიერ 1984 წელს თბილისში ჩაწერილი ხუნძური ენის აუდიომასალა, მთქმელი: ზაჰიდათ მაჰომედბეკოვა.

კავკასიური ენებისათვის არსებული ადრეული ჩანაწერების მოპოვებამ და დიგიტალიზაციამ პირველად მისცა მეცნიერებს საშუალება დიგიტალური კვლევის მეთოდები დიაქრონიულ კვლევებში გამოეყენებინათ.

1999 წლიდან იოსტ გიპერტის თანამშრომლობა ქართველ ენათმეცნიერებთან თვისობრივად ახალ ფორმას იღებს და ამიერიდან იგი სისტემატურად, საერთაშორისო სამეცნიერო პროექტების განხორციელების გზით ვითარდება. ამ პროცესის წარმატებით განხორციელება ორმა ფაქტორმა – ფინანსური და ადამიანური რესურსების გამოჩენამ განაპირობა:

1. **საერთაშორისო თანამშრომლობის დაფინანსების წყარო** – 1999 წელს გერმანული მეცნიერების განვითარების ფონდი აცხადებს კონკურსს საერთაშორისო სამეცნიერო პროგრამის **„კავკასია/შუა აზია – ერთობლივი გზები ევროპისაკენ“** განსახორციელებლად. პროგრამა მიზნად ისახავს პოსტსაბჭოთა კავკასიასა და შუა აზიაში მეცნიერების მხარდაჭერას, რათა არ მოხდეს ამ ორ რეგიონში მეცნიერების განვითარების წყვეტა;

აუდიომასალები ინტერნეტშია განთავსებული TITUS-ის პლატფორმაზე. აქვე უნდა აღინიშნოს, რომ ეს მასალები ძირითადად რეციტაციას – დასახელებული ენისათვის არსებული ტექსტური კრებულებიდან წაკითხულ მასალებს – წარმოადგენს. ტექსტებს ძირითადად შესაბამისი ენის მკვლევრები კითხულობენ და არა ამ ენის მატარებლები.

2. ფრანკფურტის უნივერსიტეტის სასწავლო პროგრამაში კავკასიოლოგიის ინტეგრირება – 2002 წელს არნ. ჩიქობავას სახ. ენათმეცნიერების ინსტიტუტის კომპიუტერული ლაბორატორიის გამგის პროფ. მანანა თანდაშვილის ფრანკფურტის უნივერსიტეტში გადასვლამ საფუძველი ჩაუყარა გერმანულ საუნივერსიტეტო სივრცეში კავკასიოლოგიის ახალი ცენტრის დაარსებასა და განვითარებას და მნიშვნელოვნად შეუწყო ხელი კავკასიოლოგიაში საერთაშორისო პროექტების ინსტიტუციონალურ განხორციელებას.

1999-2019 წლებში იოსტ გიპერტისა და მანანა თანდაშვილის ხელმძღვანელობითა და ევროპელი და ქართველი კოლეგების მონაწილეობით 15 საერთაშორისო სამეცნიერო პროექტები განხორციელდა, რომელშიც 500-მდე ქართველი და უცხოელი მეცნიერი და ახალგაზრდა მკვლევარი იღებდა მონაწილეობას და რომლის ღირებულებაც დაახლოებით 2,5 მილიონ ევროს შეადგენდა:

1. არმაზი (1999-2002): **ARMAZI**-Projekt („Kaukasische Sprachen und Kulturen: Elektronische Dokumentation") – (J. Gippert/M.Tandaschwili; Volkswagen-Stiftung, EUR 117.900);
2. საფრთხეში მყოფი კავკასიური ენები საქართველოში (2002-2005): **ECLinG**-Projekt („Bedrohte kaukasische Sprachen in Georgien") – (J. Gippert/M.Tandaschwili; Volkswagen-Stiftung, EUR 167.800);
3. საერთაშორისო კონფერენცია თემაზე „ენობრივი სიტუაცია თანამედროვე საქართველოში (12-14.2003): Internationale Konferenz „Die sprachliche Situation im gegenwärtigen Georgien"; Frankfurt a/M (J. Gippert/M.Tandaschwili; Volkswagen-Stiftung, 10.000 EUR);
4. კავკასიური წარმოშობის პალიმფსესტები (2004-2006): Kaukasische Palimpseste („Neue Wege zur wissenschaftlichen Bearbeitung von Palimpsesthandschriften kaukasischer Provenienz") – (J. Gippert, M.Tandaschwili, W. Schulze, Z. Aleksidze, J.-P- Mahé; Volkswagen-Stiftung, 155.000 EUR);
5. სოციოლინგვისტური სიტუაცია თანამედროვე საქართველოში (2005-2007): **SSGG** („Die sprachliche Situation im gegenwärtigen Georgien") – (J. Gippert/M.Tandaschwili; Volkswagen-Stiftung, EUR 189.000);
6. მათეს და მარკოზის სახარების ქართული თარგმანების კრიტიკული რედაქცია (2006-2009): Critical Edition of the Old Georgian Version of Matthew's and Mark's Gospels – Catalogue of the Manuscripts Containing the Old Georgian Translation of the Gospels – (J. Gippert, M.Tandaschwili, D. Tvaltvadze, B. Outtier; INTAS, 110.000 EUR);

7. ქართული პალიმფსესტები (2009-2012): Georgische Palimpsesthandschriften – (J. Gippert, M.Tandaschwili, D. Tvaltvadze, B. Outtier; Volkswagen-Stiftung, 99.000 EUR);

8. ფოლკსვაგენის ფონდის საერთაშორისო შემაჯამებელი კონფერენცია „შუა აზია და კავკასია მეცნიერების ფოკუსში" (17-21.2010): Internationale Konferenz "Zwischen Europa und Orient – Mittelasien/Kaukasus im Fokus der Wissenschaft" – (J. Gippert/M.Tandaschwili; Volkswagen-Stiftung, 196.000 EUR);

9. ჰინალუღურის დოკუმენტირება (2011-2014): Documentation of Khinalug – (J. Gippert, W. Schulze, M.Tandaschwili; Volkswagen-Stiftung, EUR 299.600);

10. ქართული ენის ეროვნული კორპუსი – ტექნოლოგიური ჩარჩო (2012-2015): GNC (Georgisches Nationalcorpus) – (J. Gippert/M.Tandaschwili; Volkswagen-Stiftung, EUR 390.400);

11. დრო და სივრცე ქართველურ ენებში (2012-2015): Zeit und Verortung in den Kartwelsprachen – (B. Kobalava/M.Tandaschwili; Shota Rustaveli National Science Foundation, Georgia, GEL 149.941).

12. ქართული ენის ეროვნული კორპუსი – თემატური კორპუსები (2015-2019): GNC (Georgisches Nationalcorpus – Verlängerung) – (J. Gippert/M.Tandaschwili; Volkswagen-Stiftung, EUR 60.000)

13. საერთაშორისო სადოქტორო სკოლა დიგიტალურ ჰუმანიტარიაში (2016-2017): Cross-National Doctoral School in Digital Humanities (J. Gippert/M.Tandaschwili; Volkswagen Stiftung. 88.000 EUR)

14. ბათუმის ლინგვოკულტუროლოგიური დიგიტალური არქივი (2017-2019): Batumi Linguocultural Digital Archive (BaLDAR) – (M. Tandaschwili/R. Khalvashi, Rustaweli Foundation, 150.000 GEL)

15. საერთაშორისო საზაფხულო სკოლა დიგიტალურ კავკასიოლოგიაში (2019-2020): Summerschools on digital Caucasiology (J. Gippert/M.Tandaschwili; Volkswagen Stiftung, 87.400 EUR)

იოსტ გიპერტისა და მანანა თანდაშვილის სამეცნიერო ინტერესების გაერთიანებამ, ერთი მხრივ, და საერთაშორისო და ქართული ფონდების (ფოლკსვაგენის ფონდი, ინტასი, რუსთაველის ეროვნული სამეცნიერო ფონდი) მიერ დაფინანსებული საერთაშორისო პროექტების განხორციელებამ, მეორე მხრივ, მნიშვნელოვნად დააჩქარა ქართველოლოგიის დიგიტალიზაცია და საფუძველი მოუმზადა დიგიტალური ქართველოლოგიის განვითარებას ევროპაში.

ევროპაში მიმდინარე პროცესის პარალელურად საქართველოში ქართველოლოგიის დიგიტალიზაციის პროცესი 21-ე საუკუნის დასაწყისში დაიწყო რამდენიმე სამეცნიერო და სასწავლო-აკადემიური ინსტიტუციის ბაზაზე. მათგან განსაკუთრებით აღსანიშნავია არნ. ჩიქობავას სახ. ენათმეცნიერების ინსტიტუტში პროფ. მარინა ბერიძის ხელმძღვანელობით დაწყებული მასშტაბური სამუშაო ქართული დიალექტური კორპუსის შესაქმნელად, პროფ. თინათინ მარგალიტაძის ხელმძღვანელობით წარმოებული ლექსიკოგრაფიული პროექტები და ილიას სახელმწიფო უნივერსიტეტში პროფ. ნინო დიბორჯგინიძის ხელმძღვანელობით განხორციელებული პროექტების სერია ქართული ენის კორპუსის შესაქმნელად.

ზემოთ დასახელებული, ფრანკფურტის უნივერსიტეტის, არნ. ჩიქობავას სახელობის ენათმეცნიერების ინსტიტუტის, თსუ-ში და ილიას სახელმწიფო უნივერსიტეტის მიერ განხორციელებულ სამეცნიერო პროექტებს ქვემოთ დანართის სახით წარმოვადგენთ.

1.4. დიგიტალური ქართველოლოგიის კონცეპტუალიზაციისათვის

დიგიტალური ქართველოლოგიის, როგორც ქართველოლოგიის თვისებრივად ახალი ფორმის განვითარება რამდენიმე, როგორც თეორიულ-მეთოდოლოგიური, ისე რესურსებრივი კომპონენტის არსებობას მოითხოვს. ეს კომპონენტებია:

- **ტრადიციული ქართველოლოგიის თეორიული განვითარება** (theoretical update);
- **ქართული ენის გაციფრებისათვის საჭირო ტექნოლოგიური უზრუნველყოფა**
  - ქართული ანბანის საერთაშორისო სტანდარტიზაცია და ნორმალიზაცია (UNICODE- based standards);
- **„დიდ მონაცემთა ბაზების" (big data) შექმნა:**
  - ტექსტური კორპუსების შექმნა
    - სალიტერატურო ენის კორპუსები;
    - ზეპირმეტყველების კორპუსები;
  - ელექტრონული ლექსიკონები
    - ერთენოვანი (განმარტებითი, სინონიმური, ტერმინოლოგიური, დარგობრივი) ელექტრონული ლექსიკონები;
    - ორ- და მრავალენოვანი ელექტრონული ლექსიკონები;

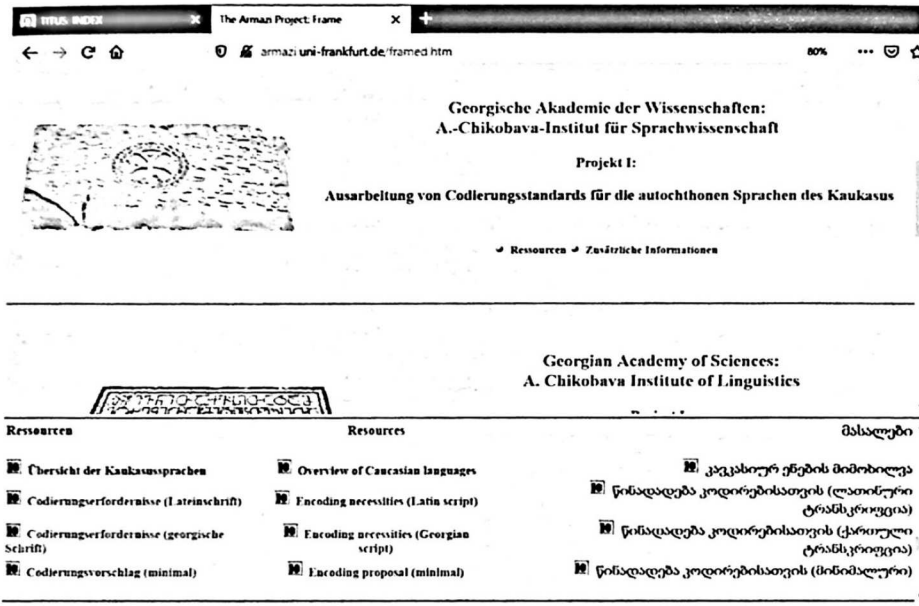
- **ქართული ენის ტექნოლოგიზაცია:**
  - ენის ავტომატური დამუშავებისათვის საჭირო ტექნოლოგიები, რომელიც ქართული ენის სისტემაზე (ფონეტიკაზე, მორფოლოგიაზე, სინტაქსზე, პრაგმატიკასა და სემანტიკაზე) იქნება მორგებული.
- **მეტაენის „გაქართულება“;**
- **ქართველოლოგთა ახალი, „დიგიტალური თაობის“ შექმნა.**

ქვემოთ შეეცდებით მოკლედ მიმოვიხილოთ დიგიტალური ქართველოლოგიის განვითარებისათვის აუცილებელი თითოეული კომპონენტი.

ქართული ენის პირველი სისტემური გრამატიკა, რომელიც აკ. შანიძეს ეკუთვნის („ქართული გრამატიკა I, მორფოლოგია“) 1930 წელს გამოიცა და ათწლეულების განმავლობაში წარმოადგენდა ქართული თეორიული აზროვნების განვითარების არსებით საფუძველს. ქართველ მეცნიერთა შემდგომი თაობის მთელი პლეადა (ივ. იმნაიშვილი, ვ. იმნაიშვილი, მზ. შანიძე, თ. გამყრელიძე, ზ. სარჯველაძე, თ. უთურგაიძე, ბ. ჯორბენაძე, დ. მელიქიშვილი, გ. გოგოლაშვილი, ა. არაბული, მ. თანდაშვილი, ნ. შარაშენიძე, ქ. დათუკიშვილი, ნ. ლოლაძე, მ. ლომია, ქ. მარგიანისუბარი და სხვ.) სწორედ აკ. შანიძის მეცნიერული თეორიის განვითარებისა და გაღრმავების გზით ავითარებს ქართულ სამეცნიერო აზროვნებას. ქართულ სამეცნიერო სივრცეში ჩნდება ენათმეცნიერების თანამედროვე მიმართულებების ადაპტაცია ქართულ ენასთან მიმართებაში, როგორიცაა მაგ.: ვალენტობა/აქტანტიანობა/პირიანობა ქართულში, მოდალობა, ევიდენციალობა, რეფერენციულობა, ქართული ზმნის ფუნქციურ-სემანტიკური მოდელები და ა.შ.). ზოგიერთი აქ ჩამოთვლილი ლინგვისტური საკითხის თეორიული დამუშავება სწორედ **დიდ მონაცემთა ბაზების** გამოყენებით გახდა შესაძლებელი.

დიგიტალური ქართველოლოგიის განვითარების მეორე აუცილებელ წინაპირობას – **ქართული ენის გაციფრების ტექნოლოგიურ უზრუნველყოფას** – ქართული ანბანის საერთაშორისო სტანდარტიზაციის და ნორმალიზაციის საკითხი წარმოადგენს, რომელზე მუშაობაც ჯერ კიდევ გასული საუკუნის 90-იან წლებში წამოიწყეს ქართველმა და გერმანელმა მეცნიერებმა.

ქართული ანბანის საერთაშორისო ნორმალიზაციისა და სტანდარტიზაციის საკითხი საერთაშორისო სამეცნიერო პროექტის „არმაზის“ ფარგლებში დამუშავდა და მასში ფრანკფურტის უნივერსიტეტის ემპირიული ენათმეცნიერების ინსტიტუტის გარდა არნ. ჩიქობავას სახ. ენათმეცნიერების ინსტიტუტისა და საქართველოს მეცნიერებათა აკადემიის თანამშრომლები იყვნენ ჩართული.



ქართველური ენების მონაცემთა გათვალისწინებით, უნიკოდში უკვე განთავსებული ქართული ანბანის მხედრული ვერსია მოგვიანებით გერმანელი ქართველოლოგის იოსტ გიპერტის ინიციატივით განიერცო რამდენიმე გრაფემით და დიაკრიტიკული ნიშნებით (10F1-10F9, 10FA, 10FB, 10FC, 10FD, 10FE, 10FF), რამაც შესაძლებელი გახადა არა მხოლოდ ქართულის, არამედ ქართველური ენების (და საქართველოში გავრცელებული წოვა-თუშურის) ნორმალიზაცია უნიკოდში[68]. დღეისათვის ქართული ანბანის სამივე სახე არის უნიკოდში ასახული: ასომთავრული (U+10A0 – U+10C5)[69], ნუსხა-ხუცური (U+2D00 – U+2D25)[70] და მხედრული (U+10D0 – U+10F5)[71].

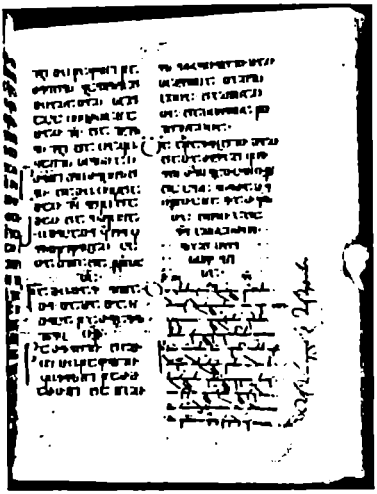
საგულისხმოა, რომ უნიკოდში ნუსხა-ხუცურის ჩამატება იოსტ გიპერტის სახელთან არის დაკავშირებული. ქართული ენის კიდევ ერთი გრაფიკული ვერსიის სტანდარტიზაციისა და ნორმალიზაციის აუცილებლობა ი. გიპერტის მიერ იმ გარე-

[68] აღნიშნული გრაფემების დამატება განხორციელდა საერთაშორისო სამეცნიერო პროექტის „ARMAZI" ერთ-ერთი ქვეპროექტის „Ausarbeitung von Codierungsstandards für die autochthonen Sprachen des Kaukasus" ფარგლებში.

[69] http://www.unicode.org/charts/PDF/U10A0.pdf

[70] http://www.unicode.org/charts/PDF/U2D00.pdf

[71] ზოგადი ინფორმაცია UNICODE-ის შესახებ შეგიძლიათ იხილოთ ვებგვერდზე http://www.unicode.org, ხოლო UNICODE 7-ის ფარგლებში შემუშავებული კოდები ვებგვერდზე http://www.unicode.org/charts/



Sin.georg. 30, fol. 75r

მოებით იქნა დასაბუთებული, რომ ზოგიერთი ქართული ხელნაწერი ქართული ანბანის სამიეე სახეს შეიცავს: ასომთაერულს, ნუსხა-ხუცურსა და მხედრულს. ასეთი ხელნაწერების რიცხვს მიეკუთვნება, მაგალითად, ერთ-ერთი ხელნაწერი სინას კოლექციიდან, კერძოდ Sin.georg.30, სადაც ასომთაერული და ნუსხური ანბანით შესრულებული ტექსტის გვერდით მხედრული ანბანით მიწერილი კოლოფონი დასტურდება. ამ ტიპის ხელნაწერების დიგიტალიზაცია კი გარდაუვლად მოითხოვს სამიეე შრიფტის უნიფიცირებას უნიკოდში.

ქართული ხელნაწერების დიგიტალიზაცია, რომელიც ჯერ კიდეე გასული საუკუნის 80-იან წლებში დაიწყო იოსტ გიპერტმა, როდესაც გრაცის უნივერსიტეტის ბიბლიოთეკაში დაცული ქართული ხელნაწერების (მათ შორის სინას ლექციონარის) დიგიტალური ვერსია შექმნა, წარმატებით გააგრძელა ქართველ კოლეგებთან ერთად საერთაშორისო პროექტების ფარგლებში:

1. 1999-2002: საერთაშორისო პროექტი „არმაზი: კაეკასიური ენები და კულტურები – ელექტრონული დოკუმენტაცია“ (ფოლკსვაგენის ფონდის დაფინანსებით, ხელმ.: ი. გიპერტი, მ. თანდაშვილი);
2. 2004-2006: კაეკასიური წარმოშობის პალიმფსესტები (ფოლკსვაგენის ფონდის დაფინანსებით, ხელმ.: ი. გიპერტი, ვ. შულცე, ზ. ალექსიძე, ჟ.-პ. მაე);
3. 2006-2009: მათეს და მარკოზის სახარების რედაქციების კრიტიკული გამოცემა (INTAS-ის დაფინანსებით, ხელმ.: ი. გიპერტი, დ. თვალთვაძე, ბ. უტიე);
4. 2009-2019: ქართული პალიმფსესტური ხელნაწერები (ფოლკსვაგენის ფონდის დაფინანსებით, ხელმ.: ი. გიპერტი, დ. თვალთვაძე, ბ. უტიე)[72];

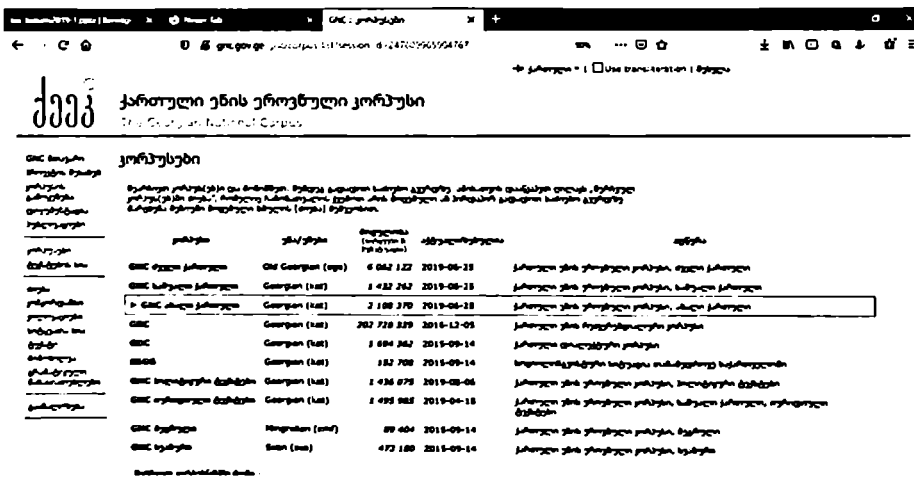
[72] პროექტი, რომელიც რამდენიმე ეტაპად ხორციელდებოდა, გასულ წელს დასრულდა, რითაც შეიქმნა ათონის ქართულ ხელნაწერთა სრული დიგიტალური კორპუსი, ათონის ქართული ხელნაწერების კატალოგი კი ქართულ, ინგლისურ და ბერძნულ ენებზე გადაცემულია დასაბეჭდად.

5. 2011-2014: ოთხთავის ქართული თარგმანების პარალელური კორპუსი (ფრანკფურტის უნივერსიტეტის დიგიტალური ჰუმანიტარიის ცენტრის LOEWE-ს პროექტი, ჰესენის მიწის მეცნიერებისა და ხელოვნების სამინისტროს დაფინანსებით, ხელმძ.: ი. გიპერტი).

შემდგომში განხორციელებული საერთაშორისო პროექტების ფარგლებში შექმნილმა რესურსებმა (როგორც ტექნოლოგიური, ისე ემპირიული თვალსაზრისით) საფუძველი მოუმზადეს წარმომადგენლობითი და დაბალანსებული სახის დიდ მონაცემთა ბაზას – **ქართული ენის ეროვნულ კორპუსს**, რომელიც ასევე საერთაშორისო სამეცნიერო პროექტების ფარგლებში შეიქმნა 2012-2019 წლებში:

6. 2012-2015: ქართული ენის ეროვნული კორპუსი – ტექნოლოგიური ჩარჩო (ფოლკსვაგენის ფონდის დაფინანსებით, ხელმ.: ი. გიპერტი, მ. თანდაშვილი);
7. 2015-2019: ქართული ენის ეროვნული კორპუსი – თემატური კორპუსები: პოლიტიკური ზეპირმეტყველების კორპუსი, იურიდიულ დოკუმენტთა კორპუსი (ფოლკსვაგენის ფონდის დაფინანსებით, ხელმ. ი. გიპერტი, მ. თანდაშვილი).

ზემოთ აღნიშნული საერთაშორისო თანამშრომლობის შედეგად შექმნილ **ქართული ენის ეროვნულ კორპუსში** გაერთიანდა ქართული ენისათვის სხვადასხვა დროს და სხვადასხვა ქვეყანაში შექმნილი დიგიტალური რესურსები: TITUS-ის, ARMAZI-ის (ფრანკფურტის უნივერსიტეტი, გერმანია), GEKKO-ს (პაულ მოირერის ქართული ენის კორპუსი, ნორვეგია) და ქართული დიალექტური კორპუსის (არნ. ჩიქობავას სახ. ენათმეცნიერების ინსტიტუტი, საქართველო) რესურსები.



ქართული ენის ეროვნულ კორპუსში ასევე გაერთიანდა ენის ელექტრონული დოკუმენტირების თანამედროვე მეთოდებით (დოკულინგვისტიკის პრინციპების გათვალისწინებით) მოპოვებული ზეპირმეტყველების მასალები, რომლებიც საერთაშორისო ერთობლივი პროექტების ფარგლებში შეიქმნა:

1. 2002-2005: საფრთხეში მყოფი კავკასიური ენები საქართველოში (ფოლკსვაგენის ფონდის დაფინანსებით, ხელმ.: ი. გიპერტი, მ. თანდაშვილი);
2. 2006-2009: სოციოლინგვისტური სიტუაცია თანამედროვე საქართველოში (ფოლკსვაგენის ფონდის დაფინანსებით, ხელმ.: ი. გიპერტი, მ. თანდაშვილი).

თუ ქართველოლოგიის განვითარებას გადავაველებთ თვალს, და შევადარებთ აქ წარმოდგენილი დიგიტალური რესურსების მოცულობასა და თემატურ მრავალფეროვნებას, ცხადად წარმოჩნდება ტრადიციული ქართველოლოგიის განვითარების გზა დიგიტალურ ქართველოლოგიამდე. კერძოდ:

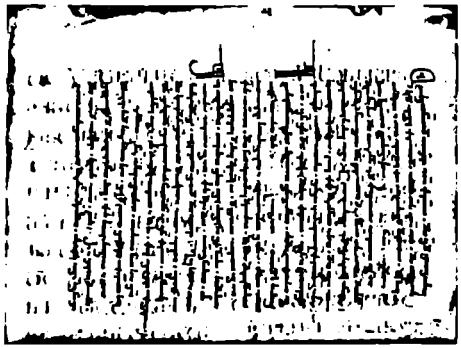
ა) ხელით დამუშავებული ტექსტებიდან დიგიტალურ კორპუსებამდე

ქართული ენის სისტემური გრამატიკის შექმნა, ისევე როგორც მრავალრიცხოვანი ნაშრომები, რომელიც ლინგვისტურ (ზოგადად ფილოლოგიურ) საკითხებს ეძღვნებოდა, მხოლოდ მრავალრიცხოვანი ტექსტების დამუშავების შედეგად იყო შესაძლებელი, რაც ხელით ხორციელდებოდა. დღეისათვის ყველა ეს ტექსტები ქართული ენის ეროვნულ კორპუსშია განთავსებული და დიგიტალური კვლევის საშუალებას იძლევა. ქართული ენის ეროვნული კორპუსის მოცულობა (დღევანდელი მონაცემებით 202 728 329 ტოკენი) საშუალებას იძლევა ქართული ენა ვიკვლიოთ როგორც სინქრონიული, ისე დიაქრონიული თვალსაზრისით.

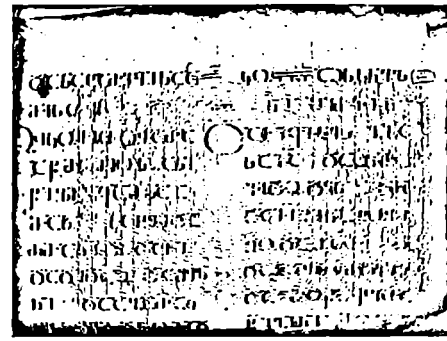
ქართული წერილობითი ძეგლების, განსაკუთრებით კი პალიმფსესტების დიგიტალიზაციამ სრულიად ახალი ბიძგი მისცა ქართველოლოგიის განვითარებას. როგორც ცნობილია, ზმნის ხანმეტი და ჰაემეტი ფორმები პირველად აკ. შანიძემ აღწერა და ემყარებოდა იე. ჯაჯანიშვილისა და აკ. შანიძის მიერ გამოცემულ ფრაგმენტებს თბილისური პალიმფსესტური ხელნაწერებიდან (A-737, A-89, A-844, H-999, H-1329 და S-3902), ასევე იაკობის პირველსახარების ვენური პალიმფსესტს Codex Vind. georg. 2, ოქსფორდულ-კემბრიჯული იერემიას ფრაგმენტს, გრაცის სინას ლექციონარს და რამდენიმე წარწერას. ძველი ქართული ენის ძეგლების დიგიტალიზაციამ, რომელიც ს'ურაბ სარჯველაძის ინიციატივით და იოსტ გიპერტთან თანამშრომლობით განხორციელდა (1997) ფრანკფურტის უნივერსიტეტში, მნიშვნელოვნად გაამდიდრა კვლევის

საბაზისო მასალა – ხანმეტური ძეგლების ნუსხა, რამაც ხანმეტი ფორმების სია მნიშვნელოვნად განაერცო[73].

პალიმფსესტების დიგიტალური დამუშავების პროცესში ტექნოლოგიური სიახლეების დანერგვამ, კერძოდ, ი. გიპერტის მიერ მულტისპექტრული აპარატის MUSIS-ის გამოყენებამ ქართული პალიმფსესტების გასაშიფრად, პრაქტიკულად 95 პროცენტამდე გაზარდა ტექსტის წაკითხვადობა. შდრ.:



ფოტოგამოსახულება



მულტისპექტრული ანალიზის შედეგი

ბ) ოთხთავის აკადემიური გამოცემიდან ოთხთავის პარალელურ კორპუსამდე

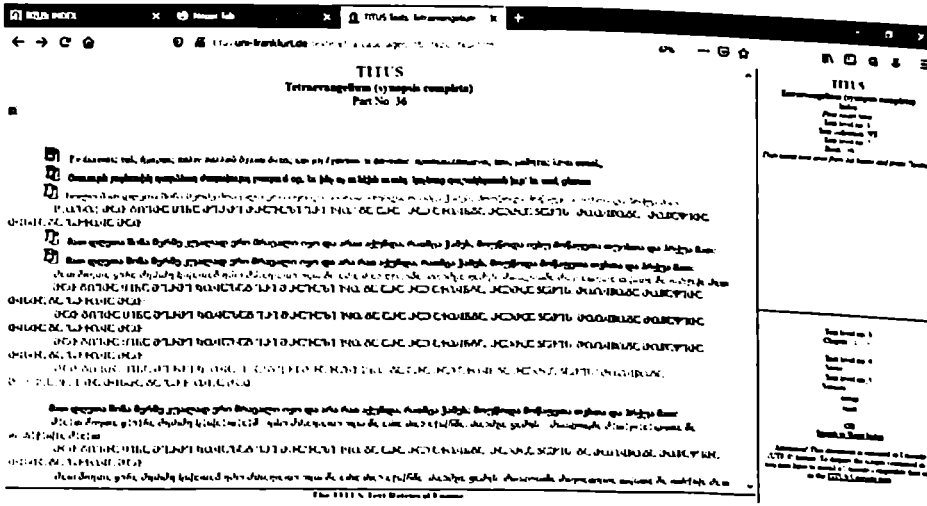
ქართველ მეცნიერებს, განსაკუთრებით კი აკ. შანიძეს, უაღრესად დიდი წვლილი მიუძღვით არა მარტო ქართული ხელნაწერების შესწავლაში, არამედ ქართული ბიბლიის ტექსტის მეცნიერულ შესწავლასა და მისი აკადემიური გამოცემის საქმეში. აკ. შანიძის მიერ 1945 წელს გამოცემული ოთხთავის აკადემიური რედაქცია, რომელიც IX-X საუკუნის ხელნაწერების მიხედვით არის დამუშავებული, უდიდესი მოვლენა იყო ქართული ფილოლოგიის ისტორიაში. ერთმანეთის გვერდით, პარალელურად წარმოდგენილი სახარების ორი რედაქცია (სამი შატბერდული ხელნაწერის მიხედვით) არსებითად პირველ პროტოტიპულ ვერსიას წარმოადგენს თანამედროვე დიგიტალური პარალელური კორპუსებისა. ივ. იმნაიშვილის ავტორობითა და აკ. შანიძის რედაქტორობით გამოცემული, ქართული ფილოლოგიის ისტორიაში ერთ-ერთი მნიშვნელოვანი სამეცნიერო ნაშრომი – ქართული ოთხთავის სიმფონია-ლექსიკონი[74] კი, რომელზედაც წლების განმავლობაში მუშაობდნენ ქართველი მეცნიერები, ტექ-

---

[73] ახლადაღმოჩენილი ხანმეტი ფორმების შესახებ იხ. ი. გიპერტი, სიახლენი ხანმეტობიდან, ენათმეცნიერების საკითხები I-II, 2009, გვ. 164-184.

[74] ი. იმნაიშვილი, ქართული ოთხთავის სიმფონია-ლექსიკონი, რედ. აკ. შანიძე, თბილისი: საქართველოს სსრ მეცნიერებათა აკადემიის გამომცემლობა, 1948-1949.

სტის დიგიტალურად დამუშავების შედეგად გენერირებული ინდექსის მანუალურად შესრულებული უსადო ვერსიაა. დიგიტალური ქართველოლოგია დღეს აქტიურად იყენებს ტექსტების დიგიტალური მეთოდებით დამუშავების ტექნოლოგიებს, მათ შორის ტექსტების დაპარალელებისა და ინდექსაციის პროგრამულ საშუალებებს. სწორედ თანამედროვე ტექნოლოგიების გამოყენებით შეიქმნა ფრანკფურტის უნივერსიტეტში **ოთხთავის ქართული თარგმანების პარალელური კორპუსი / Tetraevangelium (*synopsis completa*)**[75] საიმებო სისტემითა და სტრუქტურული ელემენტების (თავების, მუხლების) დაპარალელების შესაძლებლობით[76]. კორპუსში შესულია როგორც ხანმეტი პალიმფსესტური ხელნაწერები (A-89/A-844), ისე ჩვენამდე მოღწეული უძველესი ხელნაწერები (ადიშის (897), ჯრუჭის (H-1660), პარხლის (A-1453), ქსნის (A-509), სინური (Sin-15 და Sin-30), ანბანდიდი (Sak.ark.N-107), ოპიზის (913 წ.), ბერთას (X ს.), ტეიბერის (Xს.), ურბნისის (XI ს.), პალესტინის (1048 წ.), ქურაშის (X ს.), ვანის (1335 წ.), ეჩმიაძინის (XII-XIII ს.), გელათის (Q-908), Cod. Vind.georg.1 (XIII ს.)). **ოთხთავის ქართული თარგმანების პარალელური კორპუსი** ქართული ხელნაწერების დღეისათვის არსებულ ერთ-ერთ რეპრეზენტაციულ დიგიტალურ ბაზას წარმოადგენს.



[75] LOEWE Text collection: NT Novum Testamentum georgice, იხ. http://titus.uni-frankfurt.de/texte/etca/cauc/ageo/nt/ntkpl/ntkpl.htm

[76] პროექტი შესრულებულია ლ. სამუშიას და ნ. დუნდუას მიერ ფრანკფურტის უნივერსიტეტის დიგიტალური ჰუმანიტარიის ცენტრში, ჰესენის მიწის მეცნიერებისა და ხელოვნების სამინისტროს ფინანსური მხარდაჭერით საერთაშორისო პროექტის LOEWE ფარგლებში.

კორპუსში ოთხთაეის ქართული თარგმანები (როგორც გამოცემული, ისე გამოუცემელი) დიგიტალიზებული და დაპარალელებულია არა მარტო ცალკეული ქართული ხელნაწერების მიხედვით, არამედ შეიცავს თარგმანებისათვის რელევანტურ ბერძნულ და სომხურ დაპარალელებულ ვერსიებსაც, რაც საშუალებას აძლევს მკვლევარს მსოფლიოს სხვადასხვა გეოგრაფიულ არეალში გაბნეული ხელნაწერები სწრაფად და ეფექტიანად დაამუშაოს „ვირტუალურ სივრცეში“.

\c	**Mt.4-12**
\s	
\tl	**ვითარცა ესმა იესუს, რამეთუ იოანე მიეცა საპყრობილედ, განეშორა და წარვიდა გალილეად.**
\tg	**Ἀκούσας δὲ ὅτι Ἰωάννης παρεδόθη ἀνεχώρησεν εἰς τὴν Γαλιλαίαν.**
\ta	*ბᲐᲠᲘᲚ ᲘᲔᲡᲣᲡ ᲓᲐᲞᲧᲠᲝᲑᲘᲚᲘᲡ ᲛᲝᲡᲛᲔᲜᲐ, ᲛᲘᲐᲕᲚᲔᲜᲐ ᲘᲡᲣᲡᲛ ᲓᲐᲡᲐᲑᲐᲠᲫᲐᲜᲔᲚᲐᲓ.*
\taC	[ხ{ოლე}] ესმა რაჲ ოჳკუე ი{ესო}ჳს : რ{ამეთუ} იოანე საპყრობილედ მიეცა წარვიდა გალილეად :::
\tc	**ვითარცა ესმა იესუს, რამეთუ იოანე მიეცა საპყრობილედ, განეშორა და წარვიდა გალილეად.**
\tcR	ვითარცა ესმა იესუს, რამეთუ იოანე მიეცა, განეშორა და წარვიდა გალილეად.
\tcP	ვითარცა ესმა იესუს, რამეთუ იოანე მიეცა, განეშორა და წარვიდა გალილეად.
\tcD	ვითარცა ესმა იესუს, რამეთუ იოანე საპყრობილესა მიეცა, განეშორა და წარვიდა გალილეად.
\te	**ვითარცა ესმა იესუს, რამეთუ იოანე მიეცა საპყრობილედ, განეშორა და წარვიდა გალილეად.**
\teF	ვითარცა ესმა იესუს, რამეთუ იოანე მიეცა საპყრობილედ, განეშორა და წარვიდა გალილეად.
\teG	ვითარცა ესმა იესუს, რამეთუ ი{ოა}ნე მიეცა საპყრობილედ, განეშორა და წარვიდა გალილეად.
\teA	ესმა რაჲ ოჳკუე ი{ესო}ჳს, რ{ამეთუ} იოანე საპყრობილედ მიეცა, განეშორა და წარვიდა გალილეად.
\teB	ე{ითარც}ა ესმა ი{ესო}ჳს . რ{ამეთუ} იოანე [
\tg	**ხოლო ესმა რაჲ იესუს, ვითარმედ იოანე მიეცა, განეშორა და წარვიდა გალილეად.**
\tgH	ხოლო ესმა რაჲ იესუს, ვითარმედ იოანე მიეცა, განეშორა და წარვიდა გალილეად.
\tgI	ხოლო ესმა რაჲ იესუს, ვითარმედ იოანე მიეცა, განეშორა და წარვიდა გალილეად.
\tgK	ხოლო ესმა რაჲ იესუს, ვითარმედ იოანე მიეცა, განეშორა და წარვიდა გალილეად.
\tp	**ვითარცა ესმა იესუს, რამეთუ იოანე საპყრობილესა მიეცა, განეშორა და წარვიდა გალილეად.**

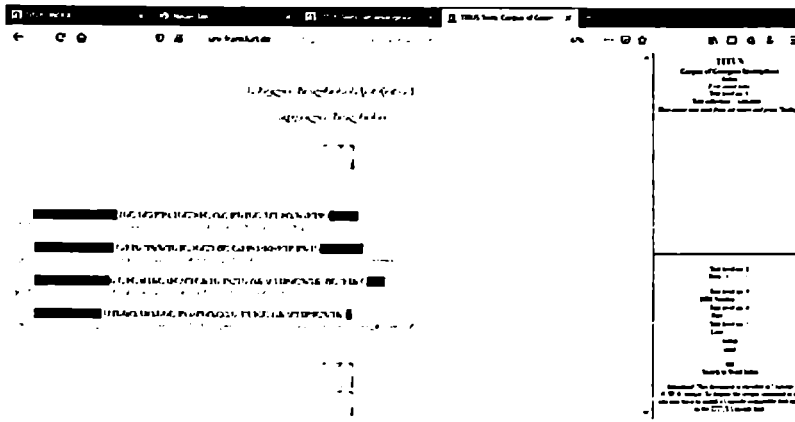
გ) წარწერების გაშიფვრიდან დიგიტალურ ეპიგრაფიკულ კორპუსებამდე

ქართული ხელნაწერების დიგიტალიზაციის პარალელურად საგანგებო პროექტების ფარგლებში განხორციელდა ქართული წარწერების დიგიტალიზაცია. დღეისათვის ორი ეპიგრაფიკული დიგიტალური კორპუსი არსებობს: ილიას სახელმწიფო უნივერსიტეტში შექმნილი „საქართველოს ეპიგრაფიკული კორპუსი“[77], რომელიც აერთიანებს საქართველოში დაცულ უმდიდრეს ეპიგრაფიკულ მემკვიდრეობას (რესურსები დოკუმენტირებულია ლაიდენის კონვენციის ციფრული მემკვიდრის – EpiDoc-ის მიხედვით) და ფრანკფურტის უნივერსიტეტში შექმნილი „ქართული წარწერების კორპუსი“[78], რომელიც აერთიანებს ქართულ წარწერებს უძველესი დროიდან

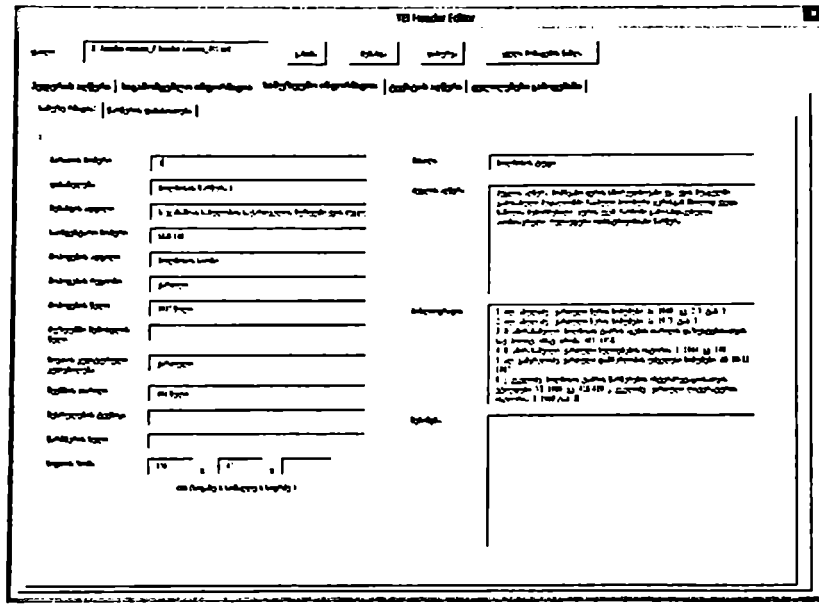
[77] https://iliauni.edu.ge/ge/iliauni/institutebi-451/lingvistur-kvlevata-centri-467/saqartvelos-epigrafikuli-korpusi

[78] http://titus.uni-frankfurt.de/texte/etcg/cauc/ageo/inscr/carcera/carce.htm

მოყოლებული დღემდე. ქართული წარწერების კორპუსი განთავსებულია TITUS-ის პორტალზე და პარალელურად ინტეგრირებულია ქართული ენის ეროვნულ კორპუსში.



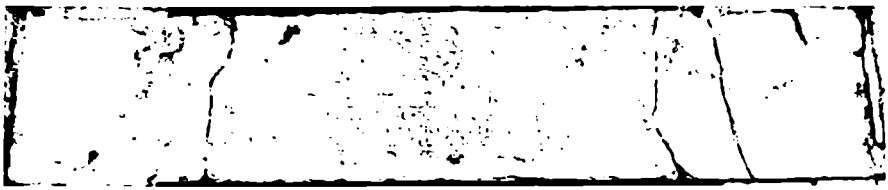
ქართული ეპიგრაფიკული ძეგლების დიგიტალიზაციამ და ელექტრონული ბაზების შექმნამ გააადვილა წარწერების სისტემატიზაცია და გააფართოვა მათი ინტერდისციპლინური კვლევის არეალი, რასაც მეტამონაცემების კომპლექსური სისტემა უზრუნველყოფს.



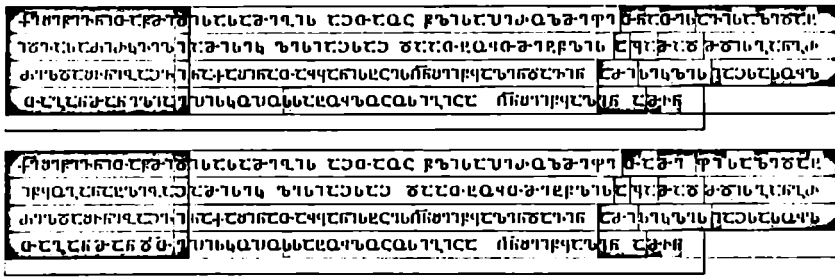
ქართული წარწერების კორპუსის მეტაინფორმაციის კომპლექსური სისტემა

როგორც ვიმოო წარმოდგენილი მეტამონაცემების კომპლექსური სისტემა გვინყენებს, წარწერების ელექტრონული ბაზები ლინგვისტური მასალის (ტექსტური რესურსი) გარდა მოიცავს ეპიგრაფიკული ძეგლის ფიზიკურ მახასიათებლებს, ძეგლის აღმოჩენის დროს, ადგილს, ლოკაციას, სამუზეუმო ექსპონატების შემთხვევაში კი საინვენტარო ნომერს. ასევე ტექსტის **იკითხვისის** ვარიანტებს და სამეცნიერო ლიტერატურის ბიბლიოგრაფიას; თანდაროული ფოტომასალა კი დამატებით ვერიფიკაციის საშუალებას აძლევს მკვლევარს ან საკითხით დაინტერესებულ პირს, თავად დარწმუნდეს დიგიტალური ვერსიის სისწორეში.

იხ. ბოლნისის სიონის წარწერა:



შდრ.: აკაკი შანიძისეული და იოსტ გიპერტისეული იკითხვისები[79]:



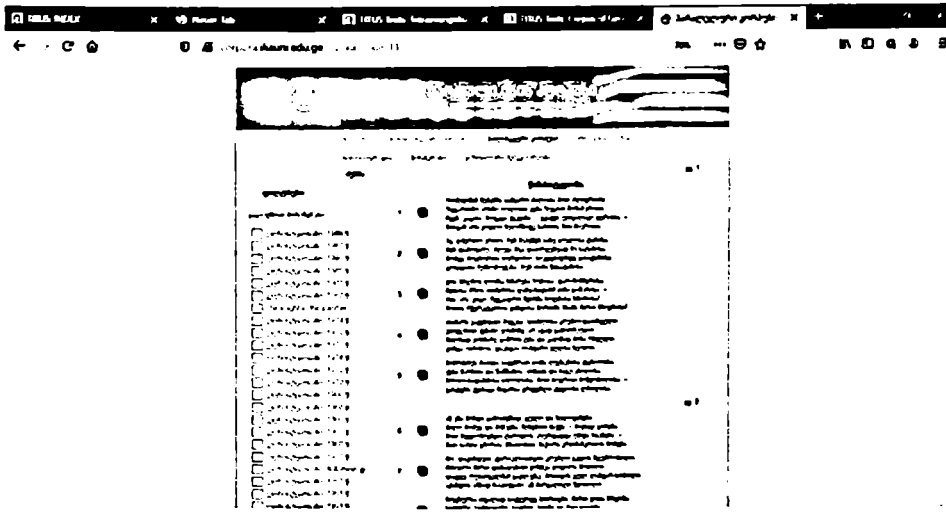
დ) ვეფხისტყაოსნის სიმფონია-ლექსიკონიდან ვეფხისტყაოსნის პარალელურ კორპუსებამდე

ცნობილია, თუ რამხელა ღვაწლი მიუძღვით ქართველ მეცნიერებს, განსაკუთრებით „ვეფხისტყაოსნის“ ტექსტის დამდგენი კომისიის წევრებს (და არა მარტო მათ) პოემის აკადემიური ტექსტის დადგენასა და გამოცემაში. გარდა ცნობილი აკადემიური გამოცემებისა[80], ცალკე გამოიცა ვეფხისტყაოსნის სიმფონია-ლექსი-

[79] J.Gippert. *Nochmals zu Beuinschrift von Bolnisi*, in: Kaukasiologie heute, Jena, Buchverlag König, 2015, 155-182 https://titus.uni-frankfurt.de/personal/jg/pdf/jg2011k.pdf

[80] იხ. შოთა რუსთაველი, ვეფხისტყაოსანი, რედ.: აკ. შანიძე, 1975 წ.; შოთა რუსთაველი, ვეფხისტყაოსანი, რედ.: ალ. ბარამიძე, კ. კეკელიძე, აკ. შანიძე, 1957 წ.

კონი[81], რაც თავისი დროისათვის უდიდესი მნიშვნელობის მქონე გახლდათ რუსთველოლოგიის განვითარებისათვის. ბუნებრივია, რომ ქართული და მსოფლიო დამწერლობის ეს უმნიშვნელოვანესი ძეგლი ქართველოლოგიის დიგიტალიზაციის პირველსავე ეტაპზე მოექცა ყურადღების ქვეშ. პოემის დიგიტალური ვერსია ჯერ კიდევ 90-იან წლებში მოამზადა იოსტ გიპერტმა ვახტანგ იმნაიშვილთან ერთად და ხელმისაწვდომი გახადა პოემა დიგიტალური კვლევების განსახორციელებლად. უკანასკნელ წლებში ილიას სახელმწიფო უნივერსიტეტში „ქართული ენის კორპუსის“ ფარგლებში განხორციელდა „ვეფხისტყაოსნის“ ხელნაწერების მასშტაბური დიგიტალიზაცია (პროექტის ხელმძ.: ნ. დობორჯგინიძე), რომელიც შოთა რუსთაველის სახელმწიფო სამეცნიერო ფონდმა დააფინანსა[82]. „ვეფხისტყაოსნის“ კორპუსში შევიდა პოემის გამოცემული (ძირითადი გამოცემები) და ხელნაწერი ნუსხების გაციფრებული ვერსიები. კორპუსი მოიცავს პოემის როგორც სრულ, ისე ფრაგმენტულ ვერსიებს და წარმოადგენს „ვეფხისტყაოსნის“ ხელნაწერთა ყველაზე სრულ ციფრულ არქივს.



2018 წელს ფრანკფურტის უნივერსიტეტში დაიწყო „ვეფხისტყაოსნის“ თარგმანების სრულმასშტაბიანი დიგიტალიზაცია (ხელმ. მ. თანდაშვილი). პროექტის მიზანია „ვეფხისტყაოსნის“ თარგმანების (55 ენაზე) პარალელური კორპუსის შექმნა, რაც საშუალებას მოგვცემს კორპუსში არა მარტო თავი მოეყაროთ თარგმანების

[81] ვეფხის-ტყაოსნის სიმფონია. თბილისი: სტალინის სახელობის თბილისის სახელმწიფო უნივერსიტეტის გამომცემლობა, 1956

[82] http://corpora.iliauni.edu.ge/?q=ka/node/11

დიგიტალურ ვერსიებს, არამედ განვახორციელოთ ორიგინალისა და თარგმანების შედარებითი ანალიზის ავტომატიზება თარგმანის სტრატეგიების კვლევისა და სხვადასხვა ენაზე პოემის თარგმანების შედარების მიზნით.



პროექტი ასევე მიზნად ისახავს თარგმანმცოდნეობის, როგორც თარგმანის პროცესის კვლევის, დიგიტალიზაციას და გულისხმობს თარგმანის სტრატეგიების კვლევის ტექნოლოგიზაციას. დღეისათვის შექმნილია ვეფხისტყაოსნის თარგმანების პარალელური კორპუსის პროტოტიპული ვერსია, რომელიც პოემის თარგმანებს შეიცავს 13 ენაზე.

ორივე ზემოთ დასახელებული კორპუსი („ვეფხისტყაოსნის“ ხელნაწერების პარალელური კორპუსი და „ვეფხისტყაოსნის“ თარგმანების პარალელური კორპუსი) წარმოადგენს „ვეფხისტყაოსნის“ ტექსტის დამდგენი კომისიის მიერ დაწყებული უდიდესი ღვაწლის კანონზომიერ და ლოგიკურ გაგრძელებას, რაც უდაოდ გახდება დიგიტალური რუსთველოლოგიის განვითარების საფუძველი.

ე) დიალექტური მასალების მოპოვებიდან დოკულინგვისტიკამდე

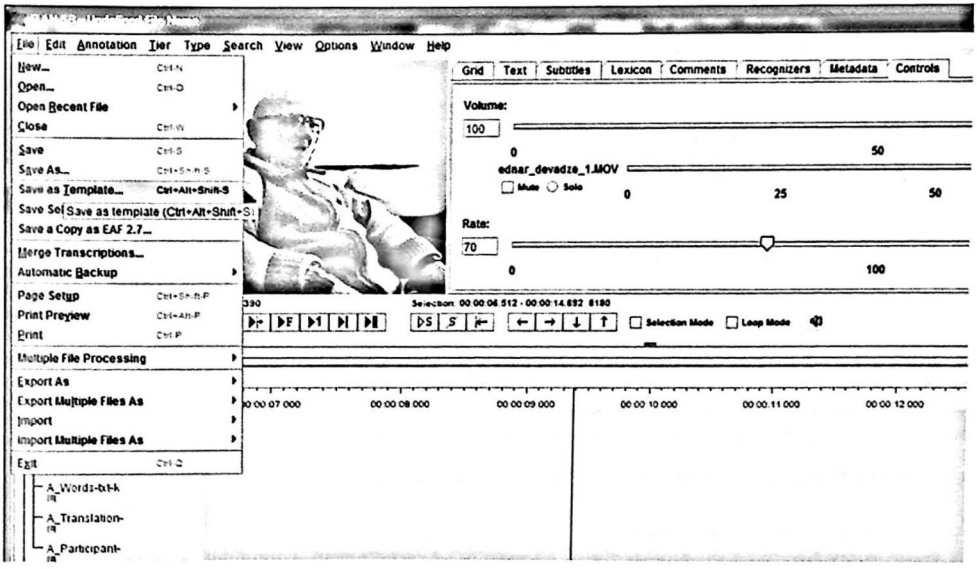
ცნობილია, რომ ქართული ენის პირველი სისტემური გრამატიკის ავტორმა აკაკი შანიძემ ქართული ენის სტრუქტურის სისტემური შესწავლა ქართული დიალექტების შესწავლით (ტექსტების ჩაწერით და მათი სისტემური ანალიზით) დაიწყო. ახალგაზრდა მეცნიერის მოღვაწეობა ამ მიმართულებით გზის გამკვლევი აღმოჩნდა დიალექტოლოგიის, როგორც სამეცნიერო დისციპლინის ჩამოყალიბებისათვის. სწო-

რედ ამიტომ, სამართლიანად არის აკ. შანიძე დიალექტოლოგიის ფუძემდებლად მიჩნეული.

21-ე საუკუნის დასაწყისში დიალექტური მასალების დიგიტალიზაცია მასშტაბურად განხორციელდა არნ. ჩიქობავას ენათმეცნიერების ინსტიტუტში, რომლის შედეგადაც შეიქმნა **„ქართული დიალექტური კორპუსი“** და რომელშიც შევიდა როგორც უკვე არსებული დიალექტური რესურსები, ისე ახლად მოპოვებული მასალები (ხელმ. მ. ბერიძე)[83].

დიალექტური მასალების მოპოვების, დიგიტალიზაციის, დაარქივებისა და შენახვის თანამედროვე მეთოდების გამოყენებით შეიქმნა მეორე „დიდ მონაცემთა ბაზა“ – **„ლინგვოკულტუროლოგიური დიგიტალური არქივი“**, რომელიც ბათუმის შოთა რუსთაველის სახელმწიფო უნივერსიტეტის დიგიტალური ჰუმანიტარიის ცენტრში განხორციელდა (ხელმ.: მ. თანდაშვილი, რ. ხალვაში)[84]. პროექტს აფინანსებდა შოთა რუსთაველის ეროვნული სამეცნიერო ფონდი.

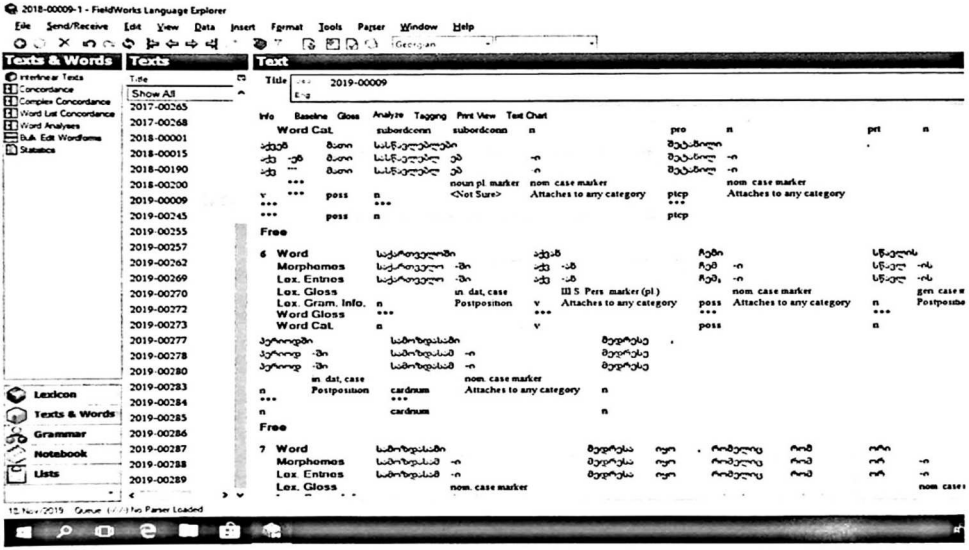
პროექტის ფარგლებში დოკუმენტირების ინოვაციური მეთოდების გამოყენებით აჭარის რეგიონში შეგროვდა ინტერდისციპლინური კვლევებისათვის განკუთვნილი ვრცელი მასალა 100 საათის მოცულობით.



[83] http://www.corpora.co

[84] http://digiarchive.bsu.edu.ge

ლინგვოკულტუროლოგიური დიგიტალური არქივის მასალები უახლესი ტექნოლოგიური პროგრამების გამოყენებით არის დამუშავებული, საარქივო მასალები დაარქივების თანამედროვე, საგანგებო ფორმატში არის დაარქივებული და შეესაბამება ნეიმეხენის მაქს პლანკ ინსტიტუტის ენათა არქივის მიერ შემუშავებულ დოკუმენტირების საერთაშორისო სტანდარტებს, კერძოდ, ვიდეო-, აუდიო- და ტექსტური მონაცემები ერთ მულტიმედიალურ ფორმატში – **ELAN Annotation Format** – (http://www.file-extension.org/de/extensions/eaf) არის გაერთიანებული, ლინგვისტური ანალიზისათვის კი გამოყენებულია ტექსტის გლოსირებისა და ანოტირების ინსტრუმენტი **Fieldworks Language Explorer (FLEx)**. იხ. არქივის მასალა წარმოდგენილი **FLEx**-ში.



ექსპედიციებში მოპოვებულ ენობრივ მონაცემთა დაარქივების როგორც ძირითადი პრინციპები და მეთოდები, ისე თანამედროვე ტექნოლოგიები, აღწერილია პროექტის ფარგლებში გამოცემულ თეორიულ ნაშრომებში: „ენისა და კულტურის დოკუმენტირების თანამედროვე მეთოდები და ინსტრუმენტები“[85] და არქივის ტექნოლოგიურ მეგზურში „ენისა და კულტურის ელექტრონული დოკუმენტირების თანამედროვე ინსტრუმენტები“[86].

---

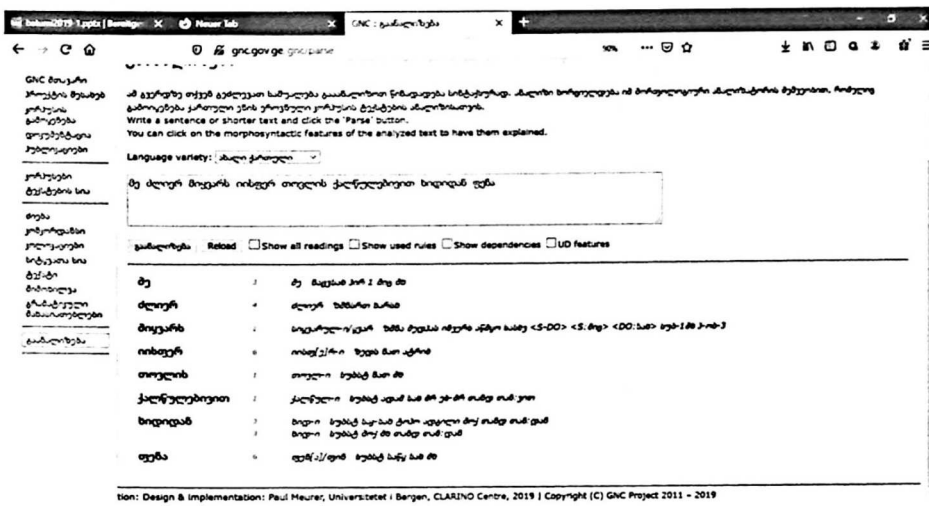
[85] იხ. მანანა თანდაშვილი, ენისა და კულტურის დოკუმენტირების თანამედროვე მეთოდები და ინსტრუმენტები“, თბილისი, გამომცემლობა უნივერსალი, 2020 წ.;

[86] მ. თანდაშვილი, რ. ხალვაში, რ. პაპიაშვილი, მ. ხახუტაიშვილი, გ. კახიანი, ნ. ცეცხლაძე, ენისა და კულტურის ელექტრონული დოკუმენტირების თანამედროვე ინსტრუმენტები (მეგზური), თბილისი, გამომცემლობა უნივერსალი, 2020 წ.

ვ) ზმნის მორფოლოგიური პარადიგმიდან ფუნქციურ-სემანტიკურ პარადიგმამდე

აკ. შანიძის მიერ სისტემატიზებული ზმნის უღლილების პარადიგმა იმ ძირითად საყრდენს წარმოადგენს, რომელზედაც აიგო შემდგომ ზმნის ახალი თეორიული თუ ფორმალური მოდელები. ეს ეხება როგორც ფორმობრივ, ისე სისტემურ მხარეს: უღლილების პარადიგმას, მწკრივთა რაოდენობას, მწკრივში ზმნურ ფორმათა რაოდენობას, ზმნის კომპლექსურ – ფუნქციურ-სემანტიკურ პარადიგმას. ყველა მომდევნო ნაშრომებში, რომელიც ქართული ზმნის ფორმალიზაციას ეძღვნებოდა, აკაკი შანიძის გრამატიკის საფუძვლები საყრდენ და ამოსავალ წერტილად რჩებოდა.

ქართული ენის დიგიტალიზაციამ თავისთავად მოითხოვა ქართული ენის გრამატიკის დიგიტალურ წესებად გადაწერის აუცილებლობაც, ე.წ. **მორფოსინტაქსური წესების დიგიტალური კორპუსის შექმნა.**



პაულ მოირერის ავტომატური ანალიზის პროგრამა ქართული ენის ეროვნულ კორპუსში

ქართული ენის ეროვნული კორპუსის სინტაქსური ანალიზატორი, რომელიც ნორვეგიელი მეცნიერის პაულ მოირერის მიერ არის შექმნილი, ერთ-ერთი ასეთი **„დიგიტალური გრამატიკის“** რეპრეზენტაციას წარმოადგენს. მიუხედავად იმისა, რომ დღეისათვის ენის ანალიზისა და სინთეზის არაერთი პროდუქტი არსებობს, პაულ მოირერის მიერ შექმნილი **მორფოსინტაქსური წესების დიგიტალური კორპუსი** იმით არის გამორჩეული, რომ იგი დიდ მონაცემთა ბაზას – **ქართული ენის ეროვნულ კორპუსს** ეყრდნობა და მასში არის ჩაშენებული.

ზ) ქართული სამეცნიერო ტერმინოლოგიის შექმნიდან დიგიტალური ჰუმანიტარიის მეტაენის გაქართულებამდე

როგორც ზემოთ უკვე აღვნიშნეთ, აკაკი შანიძის გრამატიკა ქართული ენისათვის შექმნილი პირველი სისტემური გრამატიკაა, რომელმაც საფუძველი დაუდო სისტემურ გრამატიკულ აზროვნებას და დღემდე რჩება ერთ-ერთ მნიშვნელოვან საფუძვლად დღეისათვის არსებულ ყველა გრამატიკებს შორის. გრამატიკულ ცნებათა ის სისტემა, რომელიც ქართულენოვან სამეცნიერო ლიტერატურაში გამოიყენება, ძირითადად აკაკი შანიძის შექმნილია და საყოველთაოდ გამოყენებული ქართულ ენათმეცნიერებაში. ცნობილია, თუ რა რუდუნებით ეკიდებოდა აკ. შანიძე თავის თანამოაზრეებთან ერთად ქართული სამეცნიერო ტერმინოლოგიის შექმნის საკითხს, ვინაიდან კარგად ესმოდა მეტაენის გაქართულების მნიშვნელობა ეროვნული მეცნიერების განვითარების საქმეში. სამწუხაროდ, ქართული ენის დიგიტალიზაციის პროცესი ჯერ კიდევ მოიკოჭლებს ამ თვალსაზრისით: დიგიტალიზაციას საოცარი სისწრაფით შემოაქვს ახალი ცნებები და მათი აღმნიშვნელი ტერმინები, რომელიც უმეტეს შემთხვევაში უცხოენოვანი სიტყვების ქართულ (და ხშირად არამართებულ) ტრანსლიტერაციას წარმოადგენს. დღეისათვის არსებობს დიგიტალური ჰუმანიტარიის ტერმინთა გაქართულების სულ რამდენიმე მცდელობა, მათ შორის: ა) ბეჭდური რესურსი „კორპუსლინგვისტიკის ცნებების გლოსარიუმი“[87], რომელიც დიგიტალური ჰუმანიტარიის საზაფხულო სკოლის მონაწილეებისთვის შეიქმნა 2014 წელს და რომელშიც თავმოყრილია კორპუსლინგვისტიკის ძირითადი ცნებები და შესაბამისი ტერმინების ქართული ეკვივალენტები განმარტებებითურთ და ბ) დიგიტალური რესურსი **ქართული სამეცნიერო მეტაენის კორპუსი**[88]. ამ მიმართულებით სამუშაო კიდევ ბევრია, რათა არ გაწყდეს ის ტრადიცია, რომელიც დაამკვიდრეს ქართველმა მეცნიერებმა თითქმის საუკუნის წინ ქართული სამეცნიერო ენის შესაქმნელად და მისი ეკოლოგიის შესანარჩუნებლად.

დიგიტალური ქართველოლოგიის განვითარება არ ნიშნავს მხოლოდ დიდ მონაცემთა ბაზების შექმნასა და კვლევის თანამედროვე ტექნოლოგიების განვითარებას. დარგის განვითარება სრულიად წარმოუდგენელია ადამიანური რესურსების გარეშე, კერძოდ, ახალი თაობის აღზრდისა და სამეცნიერო სივრცეში მათი ინტეგრაციის გარეშე. სამწუხაროდ, დღემდე ვერ მოხერხდა დიგიტალური ჰუმანიტარიის, როგორც სასწავლო-აკადემიური პროგრამის ან თუნდაც კურსის შეტანა საუნივერ-

---

[87] მ. თანდაშვილი, ზ. ფურცხვანიძე, კორპუსლინგვისტიკის ცნებების გლოსარიუმი, ფრანკფურტი-თბილისი, 2014

[88] http://sml.corpora.co/

სიტეტო სივრცეში. გამონაკლისს მხოლოდ რამდენიმე უნივერსიტეტი, კერძოდ ბათუმის შოთა რუსთაველის სახელმწიფო უნივერსიტეტი და ილიას სახელმწიფო უნივერსიტეტი წარმოადგენს, რომლებმაც ფართოდ გაუღეს კარი დიგიტალურ ჰუმანიტარიას.

ბათუმის შოთა რუსთაველის სახელობის სახელმწიფო უნივერსიტეტი იყო პირველი უნივერსიტეტი საქართველოში, სადაც ყველაზე ადრე დაინერგა დიგიტალური ჰუმანიტარია – 2015 წელს აკრედიტაცია გაიარა დიგიტალური ჰუმანიტარიის მაინორმა პროგრამამ (ხელმ.: პროფ. მანანა თანდაშვილი), ხოლო 2017 წელს ფრანკფურტის უნივერსიტეტისა და ბათუმის უნივერსიტეტის ურთიერთთანამშრომლობის ფარგლებში დაარსდა **დიგიტალური ჰუმანიტარიის ცენტრი** (ხელმ.: მანანა თანდაშვილი, რამაზ ხალვაში), რომელიც მიზნად ისახავს ენისა და კულტურის დოკუმენტირების თანამედროვე მეთოდების დანერგვას საქართველოში; კერძოდ, მასალების მოპოვებას დოკულინგვისტიკის პრინციპების გამოყენებით, მოპოვებული რესურსების დიგიტალურ დაარქივებას მათი ინტერდისციპლინურ კვლევებში გამოყენების მიზნით, ენათა დამუშავების თანამედროვე პროგრამული საშუალებების ELAN-ის და FLEX-ის ადაპტაციას ქართული ენისათვის. 2017-2020 წლებში ბათუმის დიგიტალური ჰუმანიტარიის ცენტრში მანანა თანდაშვილისა და რამაზ ხალვაშის ინიციატივით განხორციელდა საერთაშორისო სამეცნიერო პროექტი **„ლინგვოკულტუროლოგიური დიგიტალური არქივი“** (პროექტი დააფინანსა შოთა რუსთაველის საქართველოს ეროვნულმა სამეცნიერო ფონდმა). პროექტის ფარგლებში აჭარის რეგიონში შეგროვდა და დაარქივდა ინტერდისციპლინური ხასიათის მასალა 100 საათის მოცულობით, რომელიც ღია რესურსს წარმოადგენს და თანამედროვე ტექნოლოგიური სტანდარტების მიხედვით გამართული პირველი დიგიტალური არქივია საქართველოში.

მეორე უნივერსიტეტი, რომელმაც საგანმანათლებლო სივრცეში კარი გაუღო დიგიტალურ ჰუმანიტარიას, არის ილიას სახელმწიფო უნივერსიტეტი. 2017 წელს უნივერსიტეტში გაიხსნა სადოქტორო პროგრამა დიგიტალურ ჰუმანიტარიაში, 2020 წელს კი შეიქმნა **ლექსიკოგრაფიისა და ენობრივი ტექნოლოგიების ცენტრი** (ხელმ.: თ. მარგალიტაძე), რომლის მიზანია საქართველოში ლექსიკოგრაფიის ინტერდისციპლინურ დარგად ჩამოყალიბებისათვის ხელის შეწყობა, ქართულ ლექსიკოგრაფიაში კორპუსული და ბუნებრივი ენის დამუშავების მეთოდების ინტენსიურად დანერგვა, მაღალი ხარისხის ორენოვანი ლექსიკოგრაფიული და კორპუსული მასალის შექმნა-დაგროვება-განვითარების ხელშეწყობა, თანამედროვე ქართულ

ტერმინოლოგიაზე მუშაობა (ცენტრის მიერ შექმნილი ტერმინოლოგიური ლექსიკონების ერთ ბაზაში გაერთიანება თანამედროვე ტერმინთბანკების მეთოდოლოგიისა და ტექნოლოგიების გამოყენებით); ცენტრის ერთ-ერთ პრიორიტეტულ ამოცანას წარმოადგენს ინგლისურ-ქართული/ქართულ-ინგლისური სრულფასოვანი მანქანური თარგმნის პროგრამის შექმნა, მანქანური თარგმნის პროექტისათვის დიდი რაოდენობით მაღალი ხარისხის ინგლისურ-ქართული პარალელური წინადადებების დაგროვება, ინგლისურ-ქართული ლექსიკოგრაფიული რესურსების შეგსება-განვითარება, მათი მანქანური თარგმნის პროექტში გამოყენების მიზნით.

ქართულ საუნივერსიტეტო სივრცეში ზემოთ დასახელებულ აქტივობამდე, განსაკუთრებული როლი შეასრულა სეზონურმა სკოლებმა დიგიტალურ ჰუმანიტარიაში. კერძოდ, ბათუმის სახელმწიფო უნივერსიტეტში პროფ. მანანა თანდაშვილისა და პროფ. რამაზ ხალვაშის მიერ დაარსებულმა **ბათუმის საზაფხულო სკოლამ დიგიტალურ ჰუმანიტარიაში** და თბილისის სახელმწიფო უნივერსიტეტში პროფ. მანანა თანდაშვილისა და დარეჯან თვალთვაძის მიერ ჩატარებულმა **ზამთრის სკოლებმა დიგიტალურ ჰუმანიტარიაში**. დღემდე ჩატარებული 12 სეზონური სკოლა მიეძღვნა დიგიტალური ჰუმანიტარიის სხვადასხვა საკითხებს: დიგიტალური ჰუმანიტარია და ქართველოლოგიის სპეციალური პრობლემები, კორპუსლინგვისტიკა და ქართველოლოგიის სპეციალური პრობლემები, ქართველოლოგია და 21-ე საუკუნის გამოწვევები, თემატური კორპუსების აგების ძირითადი პრინციპები და მეთოდები, ენის ელექტრონული დოკუმენტირება, ენისა და კულტურის დოკუმენტირების ძირითადი პრინციპები და მეთოდები, დიგიტალური ქართველოლოგია, დიგიტალური კაკაისიოლოგია. აღნიშნული სკოლების პროგრამებისა და მათში მონაწილე ქართველი და უცხოელი მეცნიერების შესახებ ინფორმაცია დაწვრილებით შეგიძლიათ იხილოთ ქვემოთ (დანართი II).

წინამდებარე ქვეთავში ჩვენ შევეცადეთ წარმოგვედგინა ის უწყვეტი ხაზი, რომელიც გასულ საუკუნეში ქართველი მეცნიერების მოღვაწეობით იწყება და ჩვენ თვალწინ ყალიბდება დიგიტალურ ქართველოლოგიად. **დიგიტალური ქართველოლოგიის განვითარება კი დიგიტალური ჰუმანიტარიის ერთ-ერთი თანამედროვე, ეროვნული მიმართულების განვითარებას ნიშნავს.**

დანართი 1

ფრანკფურტის უნივერსიტეტის ემპირიული ენათმეცნიერების ინსტიტუტში ქართველურ და კავკასიურ ენათა დიგიტალიზაციის შედეგად შექმნილი რესურსები

TITUS-ის ქართული რესურსები

ა) ძველი ქართული ბიბლიური ტექსტები, აპოკრიფები და კომენტარები

1. გრაცის ხანმეტი ლექციონარი – ი. გიპერტი, ე. იმნაიშვილი, ზ. სარჯველაძე
2. ხანმეტი პალიმფსესტი A89 / A844 – ლ. ქაჯაია, ი. გიპერტი
3. ვენის ხანმეტი პალიმფსესტი – ი. გიპერტი, ზ. სარჯველაძე, ლ. ქაჯაია
4. ძველი აღთქმის ოშკისა და იერუსალიმის რედაქცია (ნაწილობრივ) – ი. გიპერტი
5. სახარების ადიშის რედაქცია – ი. გიპერტი, ე. იმნაიშვილი, ზ. სარჯველაძე, ს. სარჯველაძე, დ. თვალთვაძე
6. მცხეთის ბიბლია – ი. გიპერტი, ა. ხარანაული, ქ. დათუაშვილი, ნ. ლოლაძე, ს. სარჯველაძე, დ. თვალთვაძე, ნ. ჭუმბურიძე
7. პაელენი – ი. გიპერტი, ე. იმნაიშვილი, ს. სარჯველაძე, დ. თვალთვაძე
8. საქმენი მოციქულთანი – ი. გიპერტი, ე. იმნაიშვილი
9. კათოლიკეთა ეპისტოლენი – ი. გიპერტი, ე. იმნაიშვილი, ს. სარჯველაძე, დ. თვალთვაძე
10. იოანეს გამოცხადება – ი. გიპერტი, ე. იმნაიშვილი
11. ფსალმუნნი (A რედაქცია) – ი. გიპერტი
12. ქებათა ქებაჲ (G რედაქცია) – ზ. სარჯველაძე
13. პარიზის ლექციონარი – ი. გიპერტი
14. პერიკოპები ძველი და ახალი აღთქმიდან – ი. გიპერტი, ს სარჯველაძე, დ. თვალთვაძე
15. ახალი აღთქმის აპოკრიფები – ი. გიპერტი
    - თქუმული ღმრთისმშობელისა შობისათჳს (იაკობის პირველსახარება)
    - საკითხავი ენებისათჳს და სიკუდილისა, მოსაჴსენებული მაცხოვრისა ჩუენისა იესუ ქრისტესი, თქუმული ანანია ებრაელისაჲ (ნიკოდიმოსის სახარება)
    - წამება წმიდისა თომა მოციქულისაჲ ჰინდოეთს
16. იოანე ოქროპირის თარგმანებაჲ მათეს სახარებისაჲ – მ. შანიძე, ე. წამალაშვილი, მ. ბაბუხადია, ლ. ბასილაია, ნ. დობორჯგინიძე, დ. თვალთვაძე, თ. კარსანიძე, მ. მაჩხანელი, ნ. შარაშენიძე, ნ. შუღლაძე, თ. ცოფურაშვილი, თ. ცქიტიშვილი
17. ეეთალეს სტიქომეტრია – ს. სარჯველაძე, დ. თვალთვაძე

ბ) ჰაგიოგრაფიული და ჰომილეტიკური ტექსტები

18. წამებაჲ წმიდისა შუშანიკისი – ი. გიპერტი
19. მარტვილობაჲ და მოთმინებაჲ წმიდისა ევსტათი მცხეთელისაჲ – ი. გიპერტი
20. წამებაჲ წმიდისა და ნეტარისი მოწამისა ქრისტესისა ჰაბოსი – ი. გიპერტი
21. მოქცევაჲ ქართლისაჲ – ი. გიპერტი
22. ცხორებაჲ და წამებაჲ წმიდისა მოწამისა კოსტანტისი ქართველისაჲ – ი. გიპერტი
23. წამებაჲ წმიდისა მოწამისა გობრონისი – ი. გიპერტი
24. წამებაჲ ერმათა წმიდათაჲ რიცხჳთ ცხრათაჲ – ი. გიპერტი

25. წმიდათა ჳრმათა ორთა ძმათა დაჲთისი და ტირიჭანისი საკითხაჲი – ი. გიპერტი
26. ცხორებაჲ წმიდისა მამისა ჩუენისა იოვანე ზედაძნელისაჲ (A და B რედაქციები) – ი. გიპერტი
27. ცხორებაჲ და მოქალაქობაჲ მამისა ჩუენისა შიოსი და ევაგრესი – ი. გიპერტი
28. ცხორებაჲ და მოქალაქობაჲ მამისა ჩუენისა დავით გარესჯელისაჲ – ი. გიპერტი
29. მოქალაქობაჲ და წამებაჲ წმიდისა აბიბოს ნეკრესელ ეპისკოპოსისაჲ (A და B რედაქციები) – ი. გიპერტი
30. შრომაჲ და მოღუაწებაჲ ღირსად ცხორებისაჲ წმიდისა და ნეტარისა მამისა ჩუენისა გრიგოლი არქიმანდრიტისაჲ ხანცთისა და შატბერდისა აღმაშენებელისაჲ – ი. გიპერტი
31. ცხორებაჲ და მოქალაქობაჲ ღმერთ-შემოსილისა ნეტარისა მამისა ჩუენისა სერაპიონისი – ი. გიპერტი
32. ცხორებაჲ და მოქალაქობაჲ წმიდისა და ნეტარისა მამისა ჩუენისა ილარიონ ქართველისაჲ – ი. გიპერტი
33. ცხორებაჲ ნეტარისა მამისა ჩუენისა იოვანესი და ეფთჳმესი – ი. გიპერტი
34. ცხორებაჲ და მოქალაქობაჲ წმიდისა და ნეტარისა მამისა ჩუენისა გიორგი მთაწმიდელისაჲ – ი. გიპერტი
35. წამებაჲ წმიდისა და დიდებულისა მოწამისა არჩილისი – ი. გიპერტი
36. ცხორებაჲ და მოქალაქობაჲ წმიდისა და ნეტარისა მამისა ჩუენისა პეტრე ქართველისაჲ (A და B რედაქციები) – ი. გიპერტი
37. ცხოვრებაჲ და მოქალაქობაჲ და ღუაწლი წმიდისა და ღირსისა დედისა ჩუენისა ნინოჲსი – ი. გიპერტი
38. ცხოვრებაჲ და მოქალაქობაჲ ღირსისა და მოციქულთა სწორისა ნეტარისა ნინოჲსი – ი. გიპერტი
39. წამებაჲ და ღუაწლი წმიდათა და დიდებულთა მოწამეთა დაჲით და კონსტანტინესი – ი. გიპერტი
40. ნათარგმნი აგიოგრაფიული ტექსტები A-95 ხელნაწერიდან – ი. გიპერტი
    - წამებაჲ წმიდისა ქრისტინაჲსი – ი. გიპერტი
    - წამებაჲ წმიდისა ქრისტეფორჱსი – ი. გიპერტი, ბ. კრისტოფი
    - წამებაჲ წმიდათა კჳპრიანჱსი და იუსტინაჲსი – ზ. სარჯველაძე, ი. გიპერტი
41. სხჳა ნათარგმნი აგიოგრაფიული ტექტები – ი. გიპერტი
    - წამებაჲ წმიდათა რიფსიმესი და გაიანესი და მოჳუასთა მათთაჲ
    - ცხორებაჲ და მოქალაქობაჲ და წამებაჲ წმიდისა მღდელმოწამისაჲ გრიგოლი დიდისა სომხეთისა ეპისკოპოსისაჲ
    - ცხორებჲ და განგებაჲ წმიდისა მამისა ჩუენისა კჳრიაკოსისი
    - ცხორებაჲ და განგებაჲ წმიდისა სჳმეონ მესჳეტისაჲ
    - ცხორებაჲ და განგებაჲ წმიდისა მამისა ჩუენისა უდაბნოჲსა მთიებისა და მონაზონთა მოძღურისა ეფრემისი
    - ცხორებაჲ და მოქალაქობაჲ ნეტარისა მართაჲსი სჳმეონის დედისა
    - წამებაჲ წმიდისა მიწამისა აბდალმესისი – ე. იმნაიშვილი
    - წამებაჲ წმიდისა არეთაისი (ორი რედაქცია) – ე. იმნაიშვილი
42. ნათარგმნი აგიოგრაფიული, ჰომილეტიკური და ეგზეგეტიკური ტექსტები შატბერდის ხელნაწერიდან – ი. გიპერტი

- თქუმული წმიდისა და ნეტარისა მამისა ჩუენისა გრიგოლი ნოსელ ებისკოპოსისაჲ კაცისა შესაქმისათჳს
- თუალთაჲ ეპიფანე კჳპრელისაჲ თქუმული
- მჭეცთათჳს სახისა სიტჳეაჲ
- ეპიფანე კჳპრელი, საზომთათჳს და საწყაულთა
- დიომიდე, ფრაგმენტი დიონისიოს თრაკიელის გრამატიკის სქოლიოებიდან
- იპოლიტე რომაელი, ქრონიკონი
- იპოლიტეს განმარტებაჲ კურთხევათა მათთჳს მოსესთა ათორმეტთა მიმართ ნათესაჲთა
- წმიდისა იპოლიტეს თქუმული კურთხევათა მათთჳს იაკობისთა
- თქუმული იპოლიტესი, დაჲითისი და გოლიადისთჳს განმარტებაჲ
- ნეტარისა იპოლიტეს თქუმული თარგმანებაჲ ქებაჲ ქებათაჲ
- ნეტარისა იპოლიტეს თქუმული აღსასრულისათჳს ჟამთაჲსა, თქუმული წიგნთაგან წმიდათა ქრისტესთჳს და ანტექრისტესთჳს
- სიტჳეაჲ წმიდისა იპოლიტესი სარწმუნოებისათჳს
- თქუმული წმიდისა და ნეტარისა მამისა ჩუენისა იპოლიტესი მთავარებისკოპოსისაჲ სახჱ აღთქუმისაჲ
- მოქცევაჲ ქართლისაჲ
- ცხორებაჲ და განგებაჲ წმიდისა იაკობისი, რომელი ებისკოპოს იქმნა ქალაქსა ნასიბინს
- თარგმანებაჲ დაჲითის ფსალმუნებისაჲ
- მიწერილი დაჩისი მრჱველ ებისპოკოსისა ანტონისა

43. ნათარგმნი აგიოგრაფიული, ჰომილეტიკური და ეგზეგეტიკური ტექსტები Ath. 8 (57) ხელნაწერიდან – ი. გიპერტი
- წამებაჲ წმიდისა სტეფანჱსი პირველმოწამისაჲ
- პოვნაჲ ნაწილთაჲ წმიდისა სტეფანჱსი პირველდიაკონისა და პირველმოწამისა
- მიცვალებაჲ ნაწილთა წმიდისა სტეფანჱსთაჲ პირველდიაკონისაჲ და პირველმოწამისათაჲ იერუსალჱმით კოსტანტინეპოლის
- თქუმული გრიგოლი ხუცისა ანტიოქელისაჲ წმიდისა პირველმოწამისა სტეფანჱსთჳს
- თქუმული წმიდისა გრიგოლ ხუცისაჲ ქებაჲ პირველმოწამისა სტეფანჱსი
- ქებაჲ წმიდისა სტეფანჱსი პირველდიაკონისა და პირველმოწამისაჲ
- ჴსენებაჲ წმიდათა მოციქულთა პეტრჱსი და პაულჱსი პრომეს შინა
- წამებაჲ წმიდისა პაულჱ მოციქულისაჲ პრომს შინა
- წამებაჲ წმიდათა და ყოველად ქებულთა მოციქულთაჲ პეტრჱსი და პაულჱსი პრომს შინა
- ეპიფანე კჳპრელი, საზომთათჳს და საწყაულთა (B რედაქცია)

44. Ath. 8 (57) ხელნაწერის სხჳა ნათარგმნი აგიოგრაფიული ტექსტები – ი. გიპერტი
- საჴსენებელი წმიდისა და ღუაწლით შემოსილისა თეოდორესი ეჳქაიტელ ახლად სტრატიოტად მოყვანებულისა ცხოერებაჲ
- წამებაჲ წმიდათა მოწამეთა მესუკაევთაჲ
- წამებაჲ წმიდისა რომანოზ ახლისა მოწამისაჲ

- აღსრულებაჲ წმიდისა საჰაკისი პართევისაჲ
- წმიდისა იზიდბოზიდისი
- წამებაჲ წმიდათა მღუდელმოძღვართა სომხითისათა არისტაკესი, ორთანესი, ოსიკისი, გრიგორისი, დანიელისი
- წამებაჲ წმიდისა დედოფლისა საგდუხტისი
- სიტყვანი წამებისათჳს წმიდისა ოსკისი და მოყვასთა მისთა
- წამებაჲ წმიდისა ნერსე მთავარებისკოპოსისა სომხითისა
- წამებაჲ წმიდისა ვარდანისი და ძმათა მისთა
- წამებაჲ წმიდისა ატომისი და ძისა მისისაჲ და მოყუასთა მისთაჲ

45. ეპისტოლენი ნეტარისა ანტონისი – ი. გიპერტი
46. უდაბნოს მრავალთავი – ლ. ციხელაშვილი, ი. გიპერტი
47. ქართლის ცხოერება, ტ. I – ი. გიპერტი
48. წარტყუენვაჲ იერუსალჱმისაჲ – ი. გიპერტი
49. მოდგმაჲ კოსტანტინეპოლისი სპარსთა და სკჳთთა მიერ – ი. გიპერტი
50. ცხორებაჲ და განგებაჲ ამბა სუმეონისი, რომელსა ეწოდა ღმრთისათჳს სულელ – ი. გიპერტი, ვ. იმნაიშვილი
51. სხეა ნათარგმნი ჰომილეტიკური ტექსტები – ი. გიპერტი
    - მიცვალებაჲ ყოვლად წმიდისა და უბიწოჲსა და უმეტეს კურთხეულისა ღედოფლისა ჩუენისა მარადის ქალწულისა მარიამისი
    - საკითხაჲი ეპისტოლე მიწერილი კჳრილჱ ეპისკოპისისა იერუსალჱმელისაჲ მეფესა ღმრთისმოყუარესა
    - დღესასწაულსა წმიდათა მთავარანგელოსთასა საკითხაჲი თქუმული იოვანე მთავარეპისკოპოსისაჲ კოსტანტინეპოლელისაჲ
    - სწაელაჲ სტეფანე მონაზონისაჲ
52. ეტლთა და შჳდთა მნათობთათჳს – ქ. გოჩიტაშვილი
53. იოანე ოქროპირის ლიტურგია – ე. იმნაიშვილი, ი. გიპერტი
54. პეტრიწონის მონსტრის ტიპიკონი – ი. გიპერტი

გ) საშუალი ქართული ლიტერატურა

55. შოთა რუსთაველი, ვეფხისტყაოსანი – ი. გიპერტი, ვ. იმნაიშვილი
56. ვისრამიანი – ი. გიპერტი, თ. აბაშიძე, ხ. თოდუა
57. არჩილიანი (მეფე არჩილისი სხეა პოეტური ტექსტები) – პ. ოლივიერი, ი. გიპერტი, ა. მეფისაშვილი
58. ჩახრუხაძე, თამარიანი – ი. გიპერტი, ვ. იმნაიშვილი
59. შავთელი, აბდულმესიანი – ი. გიპერტი, ვ. იმნაიშვილი
60. მოსე ხონელი, ამირანდარეჯანიანი – ი. გიპერტი, ნ. ჭუმბურიძე

დ) ახალი ქართული ლიტერატურა და გრამატიკები

61. ბესიკი – ი. გიპერტი
62. ანტონ კათალიკოსის გრამატიკა – ლ. ქაროსანიძე, ი. გიპერტი
63. იოანე ბაგრატიონის „კალმასობა“ – ლ. ქაროსანიძე, რ. პაპიაშვილი, ი. გიპერტი
64. ზურაბ შანშოვანის გრამატიკა – თ. ლომთაძე, ი. გიპერტი

ქართველურ და კავკასიურ ენათა მასალები

65. ო. ქაჯაიას მეგრულ-ქართული ლექსიკონი – ო. ქაჯაია, ი. გიპერტი
66. სვანური პოეზია (ქართული თარგმანითურთ) – ი. გიპერტი, ნ. ნაკანი, ქ. ბაქრაძე, ტ. ფუტკარაძე
67. სვანური პროზაული ტექსტები, ტ. I (ბალსზემოური დიალექტი) – ი. გიპერტი, მ. თანდაშვილი, ი. ჩანტლაძე, რ. იოსელიანი, ქ. მარგიანი
68. სვანური ქრესტომათია – ი. გიპერტი, ი. ჩანტლაძე, ტ. ფუტკარაძე
69. უდიური სახარება – მ. თანდაშვილი, ვ. შულცე, ი. გიპერტი
70. ა. შიფნერის, ა. დირის, მ. ბეჟანოვის უდიური ტექსტები – ვ. შულცე, ი. გიპერტი
71. სამჯი დას – ვ. შულცე, ი. გიპერტი

არმაზის რესურსები

ა) ძველი ქართული ბიბლიური ტექსტები, აპოკრიფები და კომენტარები

1. სახარების წინაათონური რედაქცია – ე. გიუნაშვილი, მ. მაჩხანელი, ს. სარჯველაძე, ზ. სარჯველაძე, დ. თვალთვაძე
2. სახარების ათონური (გიორგისეული) რედაქცია – ე. გიუნაშვილი, მ. მაჩხანელი, ს. სარჯველაძე, ზ. სარჯველაძე, დ. თვალთვაძე
3. სახარების მცხეთური რედაქცია – ი. გიპერტი, ს. სარჯველაძე, დ. თვალთვაძე
4. გელათის ბიბლია (ნაწილობრივ) – ი. გიპერტი, ბ. თოფურია
5. ძველი აღთქმის რედაქცია (ნაწილობრივ) – ი. გიპერტი, ბ. თოფურია
6. ბაქარის ბიბლია (ნაწილობრივ) – ზ. სარჯველაძე
7. გელათის ბიბლია (ნაწილობრივ) – ი. გიპერტი, ბ. თოფურია
8. წიგნი რომელი დაწერა იოსებ არიმათიელმან მოწაფემან უფლისა ჩუენისა იესუ ქრისტჱსმან და თხრობაჲ აღშჱნებისათჳს ეკლესიისა რომელი იგი არს ლუდიას – ზ. სარჯველაძე, დ. თვალთვაძე
9. ანდრია კესარიელი, თარგმანებაჲ გამოცხადებისაჲ წმიდისა იოვანე მახარებელისა – რუსთაველის კომისია

ბ) ჰაგიოგრაფიული, ჰომილეტიკური და ფილოსოფიური ტექსტები

10. იოანე ზედაძნელის ცხოერება (სამი რედაქციის სინოპტიკური გამოცემა) – დ. ჩიტუნაშვილი
11. ნათარგმნი აგიოგრაფიული და ჰომილეტიკური ტექსტები (კიმენის რედაქცია) – ვ. კვანტრიშვილი
    - სიმჴნე და მოღუაწებაჲ წმიდისა მოწამისა ლუკიანოზისი
    - წამებაჲ კეთილად მძლეთა წმიდათა მოწამეთაჲ პაულე, ბილოს, თეონ, ირონ და მოყუასთა მათთაჲ
    - წამებაჲ წმიდისა მოწამისა ელიანოსისი
    - წამებაჲ წმიდისა მოწამისა ოაჲსი ქუეყანასა ბორგილთასა
    - წამებაჲ წმიდათა მოწამეთაჲ არზსი, პრომოჲსი, ელიაჲსი და სხუათა მათთანათა მოწამეთაჲ
    - წამებაჲ წმიდისა მოწამისა ლეონტისი

- საკითხავი თქუმული დღესასწაულთათუს რომელი იგი მიიწერა იერუსალემდ ისტუნიანეს მიერ ხარებისათუს და შობისა, მიგებებისათუს და ნათლისღებისა
- საკითხავი გამოცხადებაჲ წმიდათა საქარიაჲსი სუმეონისი და იაკობისი
- წამებაჲ წმიდისა ევგენიაჲსი
- საკითხავი და წამებაჲ წმიდისა იაკობ მოციქულისაჲ
- საჴსენებელსა წმიდისა სერგისსა თქუმული წმიდისა ევსები ეპისკოპოსისა
- წამებაჲ წმიდისა პროკოპისი
- ცხორებაჲ ნეტარისა ჩუენისა საბაჲსი

12. ცხორებაჲ წმიდისა და ნეტარისა და მოძღურისა ყოელისა სოფლისა და ეარსკულაესა მის განთიადისაჲ მამისა ჩუენისა იოვანე ოქროპირისაჲ, კონსტანტინეპოელელ მამადმთაერისაჲ – ი. გიპერტი, გ. შაბაშვილი, ქ. გონიტაშვილი
13. მამათა ცხოერებანი – ზ. სარჯველაძე
    - ცხორებაჲ დიდისა მამისა ჩუენისა ანტონისი
    - ცხორებაჲ წმიდისა და ნეტარისა მამისა ჩუენისა საბაჲსი
    - ცხორებაჲ და განგებაჲ წმიდისა და ნეტარისა მამისა ხარიტონისა
    - ცხორებაჲ მამისა ჩუენისა წმიდისა ეთჳმისი
    - ცხორებაჲ წმიდისა მამისა ჩუენისა თეედოსისი
    - ცხორებაჲ და მოქალაქობაჲ წმიდისა და ნეტარისა მამისა ჩუენისა ბარლაამ მოღუაწისაჲ
    - ცხორებაჲ და განგებაჲ წმიდისა მამისა ჩუენისა უდაბნოჲსა მთიებისა და მონაზონთა მოძღურისა ეფრემისი
    - ცხორებაჲ და განგებაჲ წმიდისა იოვანე ურაჰაელისა კათალიკოზისაჲ
    - ცხორებაჲ და განგებაჲ წმიდისა მამისა ჩუენისა კურიაკოზისი პალაერას
    - ცხორებაჲ და განგებაჲ წმიდათა მამათა სტეფანეწსი და ნიკონისი
    - ცხორებაჲ წმიდისა მამისა ჩუენისა პაელე ეპისკოპოსისაჲ და იოვანე ხუცისაჲ
    - ცხორებაჲ და განგებაჲ იოვანე ეპისკოპოსისა შეყენებულისაჲ
    - ცხორებაჲ და განგებაჲ ამბა სუმეონისი
    - ცხორებაჲ და განგებაჲ წმიდისა მამისა ჩუენისა ჯერასიმწსი
14. სინური მრავალთაჲი – ზ. სარჯველაძე, დ. თეალთეაძე
15. კლარჯული მრავალთაჲი – ზ. სარჯველაძე, დ. თეალთეაძე
16. ნათარგმნი აგიოგრაფიული და ჰომილეტიკური ტექსტები საკითხავი წიგნებიდან – ი. გიპერტი, დ. ჩიტუნაშვილი
    - თეოტიმოს, ცხორებაჲ და წამებაჲ წმიდისა მარინაჲსი
    - ეფრემ ასური, თქუმული წმიდისა ეფრემისი ზიარებისათუს
    - წმიდისა ეფრემისი გულისთქუმისათუს
    - ჴსენებაჲ წმიდათა კოზმან და დამიანეთაჲ, საქმენი და საკურველებანი მათნი
    - მარტჳლობაჲ წმიდათაჲ კოზმან და დამიანეთაჲ
    - ეფრემ ასური, თქუმული ეფრემისი სინანულისათუს ე
    - სწაულაჲ წმიდისა ეფრემისი მონაზონთა მიმართ მ
    - ფერის(ცვალებისათუს იესუ ქრისტჱს უფლისა ჩუენისა მთასა ზედა თაბორსა

- წამებაჲ წმიდისა ანასტასიაჲსი
- ჴსენებაჲ წმიდისა თევდორესი და წამებაჲ მისი
- ეფრემ ასური ანტისათჳს

17. მამათა სწავლანი – მ. მელიქიძე
18. ბალავარიანი – სამეცნიერო ლაბორატორია „ორიონი“
19. ალექსანდრე კეისრიელის ქრონიკა – რუსთაველის კომისია
20. გრიგოლ ნაზიანზელის საკითხაევები – რუსთაველის კომისია
21. „ელინთა ზღაპრობანი“ – ნონოსის სქოლიოები გრიგოლ ნაზიანზელის საკითხავებისა – რუსთაველის კომისია
22. (ფსევდო-)მაკარის თხზულებანი – გ. შაბაშვილი
23. ანბანურ-ანონიმური პატერიკები – რუსთაველის კომისია
24. იოანე მოსხი, ლიმონარი – ზ. სარჯველაძე
25. იაკობის ლიტურგია – ს. სარჯველაძე, ე. იმნაიშვილი
26. სასულიერო პოეზია – რუსთაველის კომისია
27. დიდი სჯულის კანონი – რუსთაველის კომისია
28. მცირე სჯულის კანონი – ქ. გონიტაშვილი
29. არსენ საფარელი, განყოფისათჳს ქართველთა და სომეხთა – ქ. გონიტაშვილი
30. იოანე ბოლნელი, ქადაგებანი – ს. სარჯველაძე
31. ეფრემ მცირეს ანდერძები – დ. თვალთვაძე
32. ამონიოს ერმისის თხზულებები – ბ. თოფურია
33. ნემესიოს ემესელი, ბუნებისათჳს კაცისა – ბ. თოფურია
34. პროკლე დიადოხოსი, პლატონურისა ფილოსოფისა კაეშირნი – ბ. თოფურია
35. იოანე პეტრიწი, განმარტებაჲ პროკლესათჳს დიადოხოსისა და პლატონურისა ფილოსოფიისათჳს – ბ. თოფურია
36. იოსიპოს წიგნი პურიისა, მოთხრობანი იუდაებრიელისა სიტყუადასაბამობისანი – ბ. თოფურია
37. ხრონოგრაფი გიორგი მონაზონისაჲ – ბ. თოფურია

გ) საშუალი ქართული ლიტერატურა

38. ქართლის ცხოერება, ტ. II – ი. გიპერტი, დ. ჩიტუნაშვილი
39. ქაიხოსრო, ომაინიანი – ქ. გონიტაშვილი
40. ვისრამიანი (მეფე არჩილის მეტრული ვერსია) – თ. აბაშიძე, ხ. თოდუა
41. ვისრამიანი (სპარსული ვერსია) – თ. აბაშიძე, ხ. თოდუა, ი. გიპერტი
42. არჩილიანი (მეფე არჩილის სხვა პოეტური ტექსტები) – პ. ოლივიერი, ი. გიპერტი, ა. მეფარიშვილი
43. შაჰნამე (ქართული ვერსიები) – ი. გიპერტი, თ. აბაშიძე, ხ. თოდუა
44. იოსებ-ზილიხანიანი – თ. აბაშიძე, ხ. თოდუა
45. თიმსარიანი – თ. აბაშიძე, ხ. თოდუა
46. რუსუდანიანი – გ. შაბაშვილი
47. შვიდი მთიები, ბარამგურიანი – გ. შაბაშვილი
48. ფეშანგი, შაჰნავასიანი – გ. შაბაშვილი
49. სულხან-საბა ორბელიანი, ქილილა და დამანა – ზ. სარჯველაძე, დ. თვალთვაძე
50. სულხან-საბა ორბელიანი, სიბრძნე სიცრუისა – ნ. დანელია

დ) მეგრული ენის ტექსტები

51. მეგრული ტექსტები კრებულიდან Сборникъ Материаловъ для описанія мѣстностей и племенъ Кавказа, вып. 10, 18) – მ. დოგრაშვილი, ე. დადიანი, ტ. ფუტკარაძე
52. მეგრული ტექსტები ი. ყიფშიძის გრამატიკიდან (ქართული თარგმანითურთ) – ე. დადიანი, ლ. კორთხონჯია, ტ. ფუტკარაძე
53. მ. ხუბუას მეგრული ტექსტები – მ. დოგრაშვილი, ე. დადიანი, ტ. ფუტკარაძე
54. კ. დანელიასა და ა. ცანავას მეგრული ტექსტები – ტ. ფუტკარაძე, მ. დოგრაშვილი, ე. დადიანი

ე) სვანური ენის ტექსტები

55. ლუშნუ ანბან – ნ. ნაკანი, ტ. ფუტკარაძე
56. სვანური ტექსტები კრებულიდან Сборникъ Материаловъ для описанія мѣстностей и племенъ Кавказа, вып. 10, 18, 31) – ი. გიპერტი, ნ. ნაკანი
57. სვანური ტექსტები კრებულიდან „ძველი საქართველო“, ტ. 2 – ნ. ნაკანი, ტ. ფუტკარაძე
58. ე. გაბლიანის სვანური ტექსტები – ნ. ნაკანი, ტ. ფუტკარაძე
59. ვ. ბარდაველიძის სვანური ტექსტები (ახალი წლის ციკლი) – ნ. ნაკანი, ტ. ფუტკარაძე
60. სვანური პროზაული ტექსტები, ტ. II (ბალსქემოური დიალექტი, ქართული თარგმანითურთ) – ნ. ნაკანი, დ. ფირცხელანი, მ. დოგრაშვილი, ტ. ფუტკარაძე
61. სვანური პროზაული ტექსტები, ტ. III (ლენტეხური დიალექტი) – მ. დოგრაშვილი, ქ. ბაქრაძე, ტ. ფუტკარაძე
62. სვანური პროზაული ტექსტები, ტ. IV (ლაშხური დიალექტი) – მ. დოგრაშვილი, ქ. ბაქრაძე, ტ. ფუტკარაძე
63. ა. ონიანის ლაშხური ტექსტები – დ. ფირცხელანი, ტ. ფუტკარაძე

ECLinG-ის რესურსები

ა) სვანური ენის რესურსები

64. კოდორის ხეობის სვანური ტექსტები – მ. საღლიანი, ქ. მარგიანი
65. სვანური ენის აუდიო-ვიდეო მასალები – რ. იოსელიანი, მ. საღლიანი, ი. ჩანტლაძე, ი. გიპერტი

ბ) უდიური ენის რესურსები

66. უდიურ–რუსულ–აზერბაიჯანული ლექსიკონი – ე. შულცე, ი. გიპერტი, ლ. ალბორნი, ლ. სამუშია, ი. ნევსკაია
67. უდიური ენის აუდიო-ვიდეო მასალები – მ. თანდაშვილი, ი. გიპერტი, გ. ტისლარიშვილი, ნ. ტისლარიშვილი, ს. ამირეჯიბი

გ) წოვა-თუშური ენის რესურსები

68. წოვა-თუშურ–ქართულ–რუსული ლექსიკონი – რ. პაპიაშვილი, ი. გიპერტი
69. წოვა-თუშური ენის აუდიო-ვიდეო მასალები – ბ. შაქელიშვილი, მ. მაჭა-ვარიანი, ს. ამირეჯიბი, ი. გიპერტი

70. აზერბაიჯანული ენის აუდიო-ვიდეო მასალები – ნ. სააკაძე, რ. რამიშვილი
71. აისორების მეტყველების აუდიო-ვიდეო მასალები – მ. ბაბუხადია
72. აფხაზური ენის აუდიო-ვიდეო მასალები – ლ. ავიძბა
73. ბეჟიტური ენის აუდიო-ვიდეო მასალები – ტ. სიხარულიძე, რ. აბაშია
74. თურქული ენის აუდიო-ვიდეო მასალები (ბერძენთა მეტყველება) – რ. პაპიაშვილი
75. ლაიჯურის აუდიო-ვიდეო მასალები – რ. რამიშვილი
76. ლაზური ენის აუდიო-ვიდეო მასალები – კ. გაბუნია, ლ. ეზუგბაია, ჭ. ქირია
77. მეგრული ენის აუდიო-ვიდეო მასალები – კ. გაბუნია, ლ. ეზუგბაია, ჭ. ქირია
78. ოსური ენის აუდიო-ვიდეო მასალები – რ. რამიშვილი, მ. ბაბუხადია
79. რუსული ენის აუდიო-ვიდეო მასალები – მ. ბაბუხადია, რ. ყაერელიშვილი და სხვ.
80. სომხური ენის აუდიო-ვიდეო მასალები – რ. ყაერელიშვილი, ნ. სააკაძე. რ. რამიშვილი
81. სვანური ენის აუდიო-ვიდეო მასალები – კ. გაბუნია, ე. დადუნაშვილი და სხვ.
82. უდიური ენის აუდიო-ვიდეო მასალები – ტ. სიხარულიძე
83. ქართული დიალექტების აუდიო-ვიდეო მასალები – თ. უთურგაიძე, მ. მარჯგალაძე, მ. ბერიძე, დ. ბუიძე, ნ. სურმავა, კ. კაკიტაძე და სხვ.
84. ქისტურის აუდიო-ვიდეო მასალები – ტ. სიხარულიძე, რ. აბაშია
85. წოვა-თუშური ენის აუდიო-ვიდეო მასალები – ტ. სიხარულიძე
86. ხუნძური ენის აუდიო-ვიდეო მასალები – ტ. სიხარულიძე, რ. აბაშია
87. ჰუნზიბური ენის აუდიო-ვიდეო მასალები – ტ. სიხარულიძე, რ. აბაშია

არნ. ჩიქობავას სახ. ენათმეცნიერების ინსტიტუტში განხორციელებული პროექტები და დიგიტალური რესურსები

1. დიდი ქართული დიალექტოლოგიური ლექსიკოგრაფიული ბაზა და ქართული დიალექტების ელექტრონული ატლასი (GEDA) (2016- 2018)
2. საქართველოს ლინგვისტური პორტრეტი 3 – ქართული დიალექტური კორპუსის მორფოლოგიური ანოტირება და დიდი დიალექტური ლექსიკოგრაფიული ბაზის შექმნა (2013 – 2015)
3. ქართული საენათმეცნიერო მეტაენის კორპუსი (სამეცნიერო ხელმძღვანელი ზაალ კიკვიძე) (2014-2017)
4. საქართველოს ლინგვისტური პორტრეტი – ქართული დიალექტური კორპუსი – II ეტაპი (სამეცნიერო ხელმძღვანელი მ. ბერიძე) (2009 -2012)
5. საქართველოს ლინგვისტური პორტრეტი – ქართული დიალექტური ტექსტების კორპუსი (სამეცნიერო ხელმძღვანელი მ. ბერიძე), 2006 – 2008
6. ქართული ენის განმარტებითი ლექსიკონი (გადამუშავებული და შევსებული ელექტრონული ვერსია)

ლინგვისტური ტექნოლოგიების ჯგუფის მიერ შექმნილი პროდუქტები:

7. ქართული ლექსიკონი (www.ganmarteba.ge)
8. ენობრივ შეცდომათა ელექტრონული ლექსიკონი (www.errata.ge)

9. ქართული ენის მორფოლოგიური პროცესორი
10. პროგრამა "ლექსიკოგრაფი"

ენის მოდელირების ასოციაციის მიერ შექმნილი პროდუქტები (www.cna.ge)

11. ორთოგრაფიული ლექსიკონის ელვერსია (ავტორები: ჭაბუკი ქორია, ვახტანგ იმნაიშვილი, ქეთევან გოჩიტაშვილი, რატი სხირტლაძე)
12. ქართული ენის გრამატიკის ელექტრონული სახელმძღვანელო
13. ქართული ენის გრამატიკის ელექტრონული სისტემა
14. ქართული ენის ელექტრონული სწავლების კურსი.

ილიას სახელმწიფო უნივერსიტეტში განხორციელებული სამეცნიერო-კვლევითი პროექტები და დიგიტალური რესურსები

1. გვიანანტიკურობისა და შუა საუკუნეების რელიგიური ისტორიოგრაფიის წყაროების კვლევა (რუსთაველის ეროვნული სამეცნიერო ფონდი, 2006-2009 წ.), ნ. დობორჯგინიძე. პროექტის შედეგები დაიბეჭდა მონოგრაფიის სახით;
2. Religioese Geschichtsschreibung der Spaetantike und des Mittelalters (AvHumboldt-Stiftung, 2006), ნ. დობორჯგინიძე, პროექტის შედეგები დაიბეჭდა მონოგრაფიის სახით;
3. მეგრულის ლინგვისტური ანალიზი (რუსთაველის ეროვნული სამეცნიერო ფონდი, 2006-2009), გ. კარტოზია, მ. ლომია, რ. გერსამია, თ. ცხადაია. პროექტის შედეგები დაიბეჭდა კრებულის სახით;
4. ანტიოქიის სამწერლობო-მთარგმნელობითი ცენტრის რეკონსტრუქციისათვის (რუსთაველის ეროვნული სამეცნიერო ფონდი, 2008-2011 წ.), ქ. ბეზარაშვილი, ნ. დობორჯგინიძე, თ. ოთხმეზური. პროექტის შედეგები დაიბეჭდა ოთხი მონოგრაფიის სახით;
5. Emanzipation der Volkssprachen im Mittelalter: Die georgische Sprache im Mittelalter. (AvHumboldt-Stiftung, Akademie der Wissenschaften zu Göttingen 2006-2009), ნ. დობორჯგინიძე. პროექტის შედეგები დაიბეჭდა გერმანულენოვანი მონოგრაფიის სახით სერიაში: Sprachen und Kulturen des christlichen Orients, Reichert-Verlag Wiesbaden;
6. იდიომური გამოთქმების ქართულ-ახალბერძნული ლექსიკონი, (შოთა რუსთაველის ეროვნული სამეცნიერო ფონდი, 2009-2010 წ.), ირინა ლობჟანიძე. პროექტის ფარგლებში მომზადდა იდიომური გამოთქმების სალექსიკონო ბაზა;
7. მეგრული ტექსტების ელექტრონული კორპუსი (რუსთაველის ეროვნული სამეცნიერო ფონდი, 2009-2012), მ. ლომია, რ. გერსამია, რ. აბაშია, ნ. მატარაძე, კ. რევაზიშვილი;
8. ქართული უესტური ენის კვლევა (USIAD-ი, Save Children International თბილისის ოფისი, საქართველოს ყრუთა კავშირი, 2011-2012), თ. მახარობლიძე. პროექტის ფარგლებში პირველად აღიწერა ქართული უესტური ენა და დაიბეჭდა უესტური ენის გრამატიკა, დაიბეჭდა სამეცნიერო და მეთოდური ხასიათის ლიტერატურა;

9. დაქტილური ანბანი (USIAD-ი, Save Children International თბილისის ოფისი, საქართველოს ყრუთა კავშირი 2012), თ. მახარობლიძე. პროექტის ფარგლებში შექმნა პირველი ქართული დაქტილური ანბანი;
10. ვეფხისტყაოსნის პარალელური (ქართულ-ინგლისური) კორპუსი, ი. ლობჟანიძე, ნ. დობორჯგინიძე, მ. საკალაშვილი, ი. გუნია, ც. ხახეიაშვილი, ნ. ბილანიშვილი, გ. შუბითიძე, ლ. კუპრეიშვილი; www.corpora.iliauni.edu.ge
11. ქართლის ცხოვრების პარალელური (ძველი ქართული-ძველი სომხური-ახალი ქართული) კორპუსი (საუნივერსიტეტო პროექტი, 2011-2012), ი. ლობჟანიძე, ს. ბერეკაშვილი, ნ. დობორჯგინიძე, ც. ხახეიაშვილი, ნ. ბილანიშვილი, გ. შუბითიძე, შ. თუმანიშვილი, ნ. დათაშვილი; www.corpora.iliauni.edu.ge
12. თანამედროვე ქართული ენის საანალიზო და საანოტაციო პროგრამები (რუსთაველის ეროვნული სამეცნიერო ფონდი, 2012-2014), ი. ლობჟანიძე, ნ. დობორჯგინიძე. პროექტის ფარგლებში შეიქმნა თანამედროვე ქართული ენის მორფოლოგიური ანალიზატორი; www.corpora.iliauni.edu.ge
13. უცნობი საქართველო: ქართულ-იტალიური წყაროები საისტორიო მეხსიერების რეკონსტრუქციისათვის (საუნივერსიტეტო პროექტი, 2013-2014), ნ. დობორჯგინიძე, თ. პატარიძე, მ. დამენია. პროექტის ფარგლებში მომზადდა და დაიბეჭდა ორტომიანი პუბლიკაცია (1. მონოგრაფია, 2. ქართული და იტალიური წყაროები);
14. XIX საუკუნის პუბლიცისტიკის კორპუსი (საუნივერსიტეტო პროექტი, 2013), ი. ლობჟანიძე, ს. ბერეკაშვილი, ნ. დობორჯგინიძე, ცენტრის ასისტენტები; www.corpora.iliauni.edu.ge
15. ილიას იერიის კორპუსი (საუნივერსიტეტო პროექტი, 2013), ი. ლობჟანიძე, ს. ბერეკაშვილი, ნ. დობორჯგინიძე, ცენტრის ასისტენტები; www.corpora.iliauni.edu.ge
16. XX საუკუნის პუბლიცისტიკის კორპუსი (საუნივერსიტეტო პროექტი, 2013), ი. ლობჟანიძე, ს. ბერეკაშვილი, ნ. დობორჯგინიძე, ცენტრის ასისტენტები; www.corpora.iliauni.edu.ge
17. ქართული სამეცნიერო ენის (სადისერტაციო ნაშრომები) კორპუსი, ი. ლობჟანიძე, ს. ბერეკაშვილი, ნ. დობორჯგინიძე, ცენტრის ასისტენტები; www.corpora.iliauni.edu.ge
18. ქართული იურიდიული კორპუსი (საუნივერსიტეტო პროექტი, 2013), ი. ლობჟანიძე, ს. ბერეკაშვილი, ნ. დობორჯგინიძე, ცენტრის ასისტენტები; www.corpora.iliauni.edu.ge
19. საბჭოური სიტყვიერების კორპუსი (საუნივერსიტეტო პროექტი, 2013), ი. ლობჟანიძე, ს. ბერეკაშვილი, ნ. დობორჯგინიძე, ცენტრის ასისტენტები; www.corpora.iliauni.edu.ge
20. თანამედროვე ქართული ენის კორპუსი (საუნივერსიტეტო პროექტი, 2013-2014), ი. ლობჟანიძე, ს. ბერეკაშვილი, ნ. დობორჯგინიძე, ცენტრის ასისტენტები; www.corpora.iliauni.edu.ge
21. ქართული ენა უცხოელთათვის: ძველი და ახალი ქართული ენის შესასწავლი მასალების ელექტრონული პორტალი (რუსთაველის ეროვნული სამეცნიერო ფონდი, 2012-2014), თ. პატარიძე, ე. ქუთათელაძე, თ. კილურაძე, ივ. ლობჟანიძე;

22. ქართული ჟესტური ენის ლექსიკონი (რუსთაველის ეროვნული სამეცნიერო ფონდი, 2013-2015), თ. მახარობლიძე, საქართველოს ყრუთა კავშირის წევრები;
23. დრო და სივრცე ქართველურ ენებში (რუსთაველის ეროვნული სამეცნიერო ფონდი, 2012-2015), მ. თანდაშვილი, ი. ქობალავა, მ. ლომია, რ. გერსამია, მოწვეული მეცნიერები;
24. ფონოსემანტიკური ლექსიკა ქართველურ ენებში. სტრუქტურული, სემანტიკური და პრაგმატიკული ასპექტები (რუსთაველის ეროვნული სამეცნიერო ფონდი, 2013-2015), რ. გერსამია, მ. ლომია, მოწვეული მეცნიერები;
25. ქართული ენის კორპუსის ომონიმიის მოხსნა და სინტაქსური მოდულისთვის სახელური ნაწილის (მსაზღვრელ-საზღვრული)წინასწარი მომზადება (საუნივერსიტეტო პროექტი, 2015);
26. საქართველოს პროსოპოგრაფია. პირველი რესპუბლიკის (1918-1921) პროსოპოგრაფიული ბაზა და სოციალური კლასტერების ანალიზი (საუნივერსიტეტო პროექტი, 2015);
27. საქართველოს ეპიგრაფიკა. V-X ს. აღმოსავლეთ საქართველოს ქართული, ბერძნული, არამეული და სომხური წარწერების კორპუსი (საუნივერსიტეტო პროექტი, 2015);
28. ჟესტური ენის ბაზა და კომპიუტერული თარგმანის მოდელი (პირველი ეტაპი) (საუნივერსიტეტო პროექტი, 2015);
29. ძველი ქართული მთარგმნელობითი კორპუსი (წინაათონური და ათონური პერიოდი)(საუნივერსიტეტო პროექტი, 2015);
30. ქართული OCR-ის ლინგვისტურ-ლექსიკური ნაწილის ამომცნობის განვითარება(საუნივერსიტეტო პროექტი, 2015);
31. იტალიაში დაცული ქართული და უცხოური წყაროების კვლევა და თარგმნა(საუნივერსიტეტო პროექტი, 2015);
32. ქართული როგორც უცხო ენა. მოსწავლის წიგნი V -დისკის ტექსტური ნაწილი და სიხშირული ლექსიკონი, სამუშაო რვეული აუდიო-დისკის ტექსტური ნაწილის და მასწავლებლის წიგნი (საუნივერსიტეტო პროექტი, 2015);
33. ეპიგრაფიკის კორპუსი (მეორე ეტაპი)(საუნივერსიტეტო პროექტი, 2016);
34. მორფოლოგიური ანალიზატორის განვითარება (მეორე ეტაპი)(საუნივერსიტეტო პროექტი, 2016);
35. საქართველოს პროსოპოგრაფიული ბაზა (მეორე ეტაპი) (საუნივერსიტეტო პროექტი, 2016);
36. V-X ს. ბილინგვური კორპუსები (პეტრე იბერის ცხოვრების ქართულ-სირიული კორპუსი, შუშანიკის წამების ქართულ-სომხური კორპუსი)(საუნივერსიტეტო პროექტი, 2016);
37. ძველი ქართული მთარგმნელობითი კორპუსი (X-XI ს.)(საუნივერსიტეტო პროექტი,2016)
38. ქართულიჟესტური ენის ბაზა და ადაპტორი(საუნივერსიტეტო პროექტი, 2016);
39. საქართველოს ეპიგრაფიკის კორპუსი (III ეტაპი)(საუნივერსიტეტო პროექტი, 2017);

40. ქართული ენის კორპუსის მორფოსინტაქსური ანოტირება (III ეტაპი)(საუნივერსიტეტო პროექტი, 2017);
41. საქართველოს პირველი რესპუბლიკის (1918-1921 წ.) პროსოპოგრაფიული ბაზა (III ეტაპი) (საუნივერსიტეტო პროექტი, 2017);
42. იდიომური გამოთქმების ონლაინლექსიკონი (რუსთაველის ეროვნული სამეცნიერო ფონდი, 2014-2017), ი. ლობჟანიძე, შ. აფრიდონიძე, ი. გუნია;
43. ქართული ჟესტური ენის ზმნის მორფოლოგია (რუსთაველის ეროვნული სამეცნიერო ფონდი, 2016-2018) – თ. მახარობლიძე;
44. მოძრაობის სემანტიკური კომპონენტები მეგრულ-ლაზურში (რუსთაველის ეროვნული სამეცნიერო ფონდი, 2016-2018), ი. ქობალაია;
45. ციფრული ჰუმანიტარია ((ციფრული ეპიგრაფიკა, კომპიუტერული და კორპუსის ენათმეცნიერება, ციფრული პროსოპოგრაფია) (რუსთაველის ეროვნული სამეცნიერო ფონდი, 2017-2020), ნ.დობორჯგინიძე;
46. ქართულ-ინგლისური ონლაინლექსიკონი (I ეტაპი);
47. ძველი ქართული ენის კორპუსი და მორფოლოგიური ანალიზატორი (I ეტაპი);
48. საქართველოს პროსოპოგრაფიული ბაზა (IV ეტაპი);
49. LaDyCa, ილიას სახელმწიფო უნივერსიტეტისა (თ. მახარობლიძე) და პარიზის სორბონას უნივერიტეტის ერთობლივი პროექტი, ხელმძღვანელი პროფ. ჟან ლეო ლეონარდი;
50. ჟესტური ენის სოციოლინგვისტური კვლევა. აშშ სან ბერნანდინოს უნივერსიტეტისა და ილიაუნის (თ. მახარობლიძე) ერთობლივი პროექტი, ხელმძღვანელი პროფ. ჟენდი სმიდტი;
51. ფუნდამენტური, საზღვარგარეთ დაცული მატერიალური ძეგლების შესწავლა ნ. დობორჯგინიძის ხელმძღვანელობით (რუსთაველის ეროვნული სამეცნიერო ფონდის დაფინანსებით);
52. საქართველოს ეპიგრაფიკული კორპუსი, ნ. დობორჯგინიძის ხელმძღვანელობით (რუსთაველის ეროვნული სამეცნიერო ფონდის დაფინანსებით);
53. ძველი ქართული ენის კორპუსული ლექსიკონი (წინაათონური პერიოდი);
54. საქართველოს პროსოპოგრაფია (ქართული, უცხოური სერიული დოკუმენტები), V ეტაპი;
55. ქართული ენის კორპუსი, ციფრული რესურსები (ძველი ქართული ელინოფილური პერიოდის ბაზა, ანალიზატორი);
56. საქართველოს ეპიგრაფიკის დოკუმენტირების სამარჯვენი, II ეტაპი.

საქართველოს მეცნიერებათა ეროვნულ აკადემიაში შექმნილი დიგიტალური რესურსები

ქართული ენის თესაურუსი (მოიცავს ბიბლიისა და „ვეფხისტყაოსნის“ კორპუსებს) (http://cna.giasher.com/index.php)

თბილისის სახელმწიფო უნივერსიტეტის ლექსიკოგრაფიულ ცენტრში შექმნილი დიგიტალური რესურსები

1. დიდი ქართულ-ინგლისური ლექსიკონი (www.dict.ge)
2. ინგლისურ-ქართული სამხედრო ონლაინლექსიკონი (http://mil.dict.ge)
3. ინგლისურ-ქართული ბიოლოგიური ონლაინლექსიკონი" (http://bio.dict.ge)
4. ინგლისურ-რუსულ-ქართული ტექნიკური ტერმინოლოგიის ონლაინლექსიკონი (https://techdict.ge/)
5. ინგლისურ-ქართული პარალელური კორპუსი (სამეცნიერო ტექსტებისა და მხატვრული ლიტერატურის ქვეკორპუსებით)

თავი მესამე

კორპუსლინგვისტიკის არსი და მნიშვნელობა ქართველოლოგიის დიგიტალიზაციისათვის

1. კორპუსული მოძრაობის ისტორია საქართველოში და დიდ მონაცემთა ბაზები

ქართველოლოგიის დიგიტალიზაციის პროცესის განვითარებისათვის მნიშვნელოვანი როლი შეასრულა ქართველოლოგიაში დაწყებულმა კორპუსულმა მოძრაობამ. კორპუსლინგვისტიკის შემოტანა ქართველოლოგიაში ორ ფაქტორთან იყო დაკავშირებული: ა) ქართული ენის დიგიტალური ბაზების შექმნასთან, რომელმაც პრაქტიკული საფუძველი შეუქმნა კორპუსულ მოძრაობას ქართველოლოგიაში და ბ) ქართულ სამეცნიერო აზროვნებაში ენის კორპუსლინგვისტური პარადიგმის შემოტანასთან, რომელმაც ენის კოგნიტური და ფუნქციური გრამატიკის საკვანძო საკითხები წამოწია წინა პლანზე და მნიშვნელოვანი ბიძგი მისცა ქართველოლოგიის თეორიულ განვითარებას. ქვემოთ მოკლედ მიმოვიხილავთ თითოეულს.

ქართული ენისათვის კორპუსების შექმნა სხვადასხვა დროსა და სხვადასხვა გეოგრაფიულ გარემოში დაიწყო, რამაც მნიშვნელოვნად განაპირობა როგორც კორპუსების სტრუქტურა, ისე თემატიკა. კორპუსული მოძრაობა ქართველოლოგიაში დაკავშირებულია ოთხი მეცნიერის – **იოსტ გიპერტის, მანანა თანდაშვილის, პაულ მოირერისა და მარინა ბერიძის** სახელთან. საგულისხმოა, რომ სწორედ მათი ინიციატივით წამოწყებულმა კორპუსულმა აქტივობებმა მოუმზადეს საფუძველი ქართული ენისათვის დიდ მონაცემთა ბაზის – **ქართული ენის ეროვნული კორპუსის** შექმნას.

ქართული ენისათვის კორპუსების შექმნის პირველ ეტაპზე კორპუსები თემატურად ასე გადანაწილდა: ძველი და საშუალი ქართულის კორპუსების შექმნა დაიწყო გასული საუკუნის 80-იან წლებში გერმანიაში, იოსტ გიპერტის მიერ, რის შედეგადაც TITUS-ის პორტალზე გაჩნდა ძველი და საშუალი ქართული ენის წერილობითი ძეგლების დიაქრონიული კორპუსის პირველი პროტოტიპული ვერსია[89]. თანამედროვე ქართული ენის კორპუსის შექმნა კი გასული საუკუნის მიწურულს დაიწყო, კერძოდ ნორვეგიაში, როდესაც ინფორმატიკის სპეციალისტმა და ლინგვისტმა პაულ

[89] იხ. http://titus.uni-frankfurt.de/indexd.htm

მოიერმა თავი მოუყარა ინტერნეტში განთავსებულ ქართული ენის რესურსებს და საფუძველი ჩაუყარა თანამედროვე ქართული ენის კორპუსს[90].

საქართველოში კორპუსლინგვისტიკა საწყის ეტაპზე პრაქტიკულად იატაკქვეშეთში ვითარდებოდა და მთელი ათწლეული დასჭირდა მის ლეგიტიმაციას ქართულ სამეცნიერო სივრცეში. პირველი ქართული კორპუსლინგვისტური რესურსი – **ქართული ზეპირმეტყველების კორპუსი** – მარინა ბერიძის მიერ შეიქმნა და **ქართული დიალექტური კორპუსის** პროტოტიპულ ვერსიას წარმოადგენდა.

სამივე აქ დასახელებული კორპუსის შექმნა ერთმანეთისაგან დამოუკიდებლად განხორციელდა და თითოეულ მათგანს განსხვავებული წინაპირობები გააჩნდა. იოსტ გიპერტის მიერ გასული საუკუნის 80-იან წლებში დაწყებული ქართული ენის რესურსების დიგიტალიზაცია ძირითადად ძველი ქართულის ძეგლების გაციფრებას ისახავდა მიზნად, რომელთა ნაწილი ბეჭდურად უკვე იყო გამოცემული. მოგვიანებით, არმაზის პროექტის ფარგლებში მოხდა მისი თემატური გაფართოება – ძველი ქართულის ძეგლებს დაემატა საშუალი ქართულის ძეგლების დიგიტალური ვერსიები. საგულისხმოა, რომ გიპერტის მიერ ჩაფიქრებული იდეის განხორციელება – ძველი და საშუალი ქართულის კორპუსების ბაზაზე დიაქრონიული კორპუსის შექმნა – დიდ ძალისხმევასთან იყო დაკავშირებული. მართალია რესურსების დიდი ნაწილი უკვე არსებობდა ბეჭდური სახით და მათი დიგიტალიზაცია ტექნიკური საკითხების გადაჭრასთან იყო დაკავშირებული (სკანირებული ტექსტების მანქანური „წაკითხვა“ და ამგვარად „წაკითხული“ ტექსტების რედაქტირება), მაგრამ ძველი ქართული ენის უამრავი რესურსი, განსაკუთრებით საქართველოს გარეთ დაცული (მათ შორის პალიმფსესტური) ხელნაწერები, ჯერ კიდევ არ იყო გამოცემული და მათი დიგიტალიზაცია მთელ რიგ პრობლემებთან იყო დაკავშირებული, მათ შორის, ქართული ანბანისათვის თანამედროვე ტექნოლოგიების ადაპტაციასთან[91]. ამ მასშტაბის ამოცანების განსახორციელებლად, გარდა სოლიდური დაფინანსების მოპოვებისა, საჭირო იყო

---

[90] იხ. http://clarino.uib.no/korpuskel/corpus-list?session-id=249325642105882

[91] ბეჭდურად გამოცემული ძველი ქართული ძეგლები მხედრული ანბანით იყო დაბეჭდილი. მათი დიგიტალიზაცია კი, გარდა ტექსტის „გაციფრებისა“, მოითხოვდა ხელნაწერის ორიგინალის სრულად ადეკვატური ვერსიის შექმნას, ანუ მის გადაყვანას ასომთავრულზე ან ნუსხა-ხუცურზე, რომლის ელექტრონული შრიფტები იმ დროისათვის არ არსებობდა და რომლის შექმნა იოსტ გიპერტს თავად მოუხდა. უნიკოდში ქართული შრიფტების საერთაშორისო სტანდარტიზაციის განხორციელებამდე გიპერტის მიერ შექმნილი შრიფტი TITUS Siberbit ფუნქციონირებდა. აღნიშნული შრიფტი იხმარებოდა აგრეთვე ქართული ტექსტური ბაზებისათვის, რომლებიც WordCruncher-ში ინტეგრირებული.

ხელნაწერთა დიგიტალიზაციის ნებართვის აღება, საერთაშორისო აკადემიური რესურსების კონსოლიდაცია და ა.შ. გარდა ამისა, აღნიშნული დიაქრონიული კორპუსის შექმნა ქართული ენისათვის კორპუსლინგვისტური მიდგომის განვითარებასაც მოითხოვდა (ქართულ ენაზე მორგებული საძიებო სისტემის შექმნა), რაც იოსტ გიპერტმა მომდევნო პროექტებში („კავკასიური წარმოშობის პალიმფსესტების დიგიტალიზაცია“ და „ქართული პალიმფსესტების დიგიტალიზაცია“) წარმატებით განახორციელა ქართველ და უცხოელ მეცნიერებთან ერთად ფოლკსვაგენის ფონდის ფინანსური მხარდაჭერით. აღნიშნული პროექტის ფარგლებში იოსტ გიპერტის ინიციატივით პალიმფსესტების დასამუშავებლად პირველად იქნა გამოყენებული მულტისპექტრული აპარატი MUSIS-ი, რამაც საგრძნობლად გააადვილა ხელნაწერების „კითხვადობა“ და ეფექტიანი გახადა ტექსტის დადგენისა და კვლევის პროცესი.

დიაქრონიული კორპუსის შესაქმნელად, გარდა ძველი და საშუალი ქართული ენის ძეგლებისა, აუცილებელი იყო თანამედროვე ქართული ენის კორპუსის შექმნა, რომლის იდეა და პრაქტიკული განხორციელება ეკუთვნის ნორვეგიაში მოღვაწე კორპუსლინგვისტს პაულ მოირერს, რომელიც 90-იან წლებში დაინტერესდა ქართული ენით და საუკუნის დასაწყისში შექმნა თანამედროვე ქართული ენის კორპუსი GEKKO, რომელიც დამატებით აღჭურვილი იყო ქართული ენის მორფოლოგიური ანალიზატორით. პაულ მოირერის მიერ წამოწყებული სამუშაო შედარებით სწრაფად განხორციელებადი იყო იმდენად, რამდენადაც ქართული ენის ციფრული მასალების მოპოვება საუკუნეთა მიჯნაზე უკვე აღარ წარმოადგენდა სირთულეს და მის მიერ შექმნილი კორპუსის მასალები ძირითადად ინტერნეტსივრცეში არსებული პორტალების – lib.ge და opentext – რესურსებისაგან შედგებოდა. ორიენ აქ დასახელებული კორპუსი წერილობითი ენის რესურსებისაგან შედგებოდა და ნორმირებული ენის სისტემას ასახავდა. ქართული ზეპირმეტყველების ამსახველი კორპუსები ცოტა მოგვიანებით შეიქმნა. 2006 წელს არნ. ჩიქობავას სახ. ენათმეცნიერების ინსტიტუტში მარინა ბერიძის ინიციატივით დაიწყო მასშტაბური პროექტი, რომელიც მიზნად ისახავდა, ერთი მხრივ, ქართული დიალექტების არსებული რესურსების გაციფრებას, რომელიც არნ. ჩიქობავას სახ. ენათმეცნიერების ინსტიტუტის არქივში ინახებოდა, და, მეორე მხრივ, ზეპირმეტყველების ახალი ნიმუშების ჩაწერასა და მათ დიგიტალიზაციას. პროექტის განხორციელების შედეგად შეიქმნა ზეპირმეტყველების პირველი კორპუსი, რომელმაც საფუძველი ჩაუყარა პირველ კორპუსს ქართულ სამეცნიერო სივრცეში, კერძოდ **ქართულ დიალექტურ კორპუსს.**

ქართველური და კავკასიური ენების ზეპირმეტყველების დოკუმენტირება პარალელურად ევროპაშიც დაიწყო. 2003-06 წლებში იოსტ გიპერტისა და მანანა თანდაშვილის ხელმძღვანელობით განხორციელდა საერთაშორისო სამეცნიერო პროექტი, რომელიც საქართველოში გავრცელებული და საფრთხეში მყოფი კავკასიური ენების დოკუმენტირებას ისახავდა მიზნად. აღნიშნული პროექტის განხორციელებას გადამწყვეტი მნიშვნელობა ჰქონდა კავკასიოლოგიის დიგიტალიზაციისათვის როგორც თემის, ისე მასალების მოპოვებისა და დამუშავების ტექნოლოგიური საშუალებების გამოყენების თვალსაზრისით. კერძოდ:

1. ქართულ სამეცნიერო სივრცეში პირველად შედგა საკვლევი თემის **„საფრთხეში მყოფი ენები"** შემოტანა და მისი მეცნიერული დამუშავება კავკასიურ ენებთან მიმართებაში;
2. ახალი რესურსების მოპოვება განხორციელდა **დოკულინგვისტიკის** პრინციპებისა და მეთოდების გათვალისწინებით;
3. რესურსების ტექნოლოგიური დამუშავება განხორციელდა ინოვაციური ტექნოლოგიების, კერძოდ, **მულტიმედიალური ფორმატის EAF**-ის გამოყენებით, რომელიც საგანგებოდ შეიქმნა საფრთხეში მყოფის ენების დოკუმენტირების პროგრამის **DOBES** (Dokumentation bedrohter Sprachen) ფარგლებში (პროგრამას აფინანსებდა ფოლკსვაგენის ფონდი);
4. პროექტის შედეგად მოპოვებული და დამუშავებული რესურსები ინტეგრირებულ იქნა ნეიმეხენის მაქს პლანკის ინსტიტუტის არქივში **TLA** (The Language Archive), რითაც პირველად განხორციელდა ქართველური და კავკასიური ენების მასშტაბური ინტეგრაცია საერთაშორისო სამეცნიერო სივრცეში.

აღნიშნული პროექტის შემდგომ კიდევ ერთი პროექტი განახორციელეს იოსტ გიპერტმა და მანანა თანდაშვილმა ზეპირმეტყველების კორპუსის შესაქმნელად, რომელიც ამჯერად უკვე საქართველოში გავრცელებული ენების დოკუმენტირებას ისახავდა მიზნად. პროექტი სოციოლინგვისტური ხასიათის იყო („სოციოლინგვისტური სიტუაცია თანამედროვე საქართველოში") და გარდა საქართველოში გავრცელებული ენების ზეპირმეტყველების ნიმუშების შეკრებისა და გაციფრებისა, მიზნად ისახავდა ენობრივი სიტუაციის გამოკვლევას საქართველოში და მიღებული შედეგების თეორიულ განზოგადებას[92].

---

[92] აღნიშნულ პროექტში მიღებული შედეგები თეორიულად დამუშავდა და გამოიცა პროექტში მონაწილე მეცნიერების გ. გოგოლაშვილის, ტ. სიხარულიძის, კ. გაბუნიას, ლ. ესუგბაიას, ჯ. ქირიას მიერ.

ზემოთ აღწერილი კორპუსების შექმნამ მყარი ნიადაგი მოამზადა ქართული ენისათვის დიდ მონაცემთა ბაზის შესაქმნელად. 2012 წელს იოსტ გიპერტმა, მანანა თანდაშვილმა და პაულ მოირერმა შეიმუშავეს **ქართული ენის ეროვნული კორპუსის** შექმნის კონცეფცია, რომელიც სამეცნიერო პროექტის სახით განხორციელდა 2012-2019 წლებში. პროექტი, რომელსაც ამჯერადაც ფოლკსვაგენის ფონდი აფინანსებდა, მიზნად ისახავდა ქართული ენისათვის უკვე არსებული კორპუსების გაერთიანებას, რომელშიც გარდა წერილობითი ძეგლებისა, სპირმეტყველებაც იქნებოდა ასახული. საინიციატივო სამეცნიერო ჯგუფს შესაბამისად შემოუერთდა მარინა ბერიძე, რამაც შესაძლებელი გახადა ქართული დიალექტური კორპუსის მასალების ეროვნულ კორპუსში ინტეგრირება. ქართველ და უცხოელ მეცნიერთა თანამშრომლობის შედეგად შეიქმნა **ქართული ენის ეროვნული კორპუსის პროტოტიპული ვერსია**, რომელშიც გაერთიანდა უკვე არსებული დიგიტალური რესურსები, შემუშავდა კორპუსის ტექნოლოგიური ჩარჩოს „ოქროს სტანდარტი“ და კორპუსის შემდგომი განვითარების – **ეტაპობრივი დაბალანსების** – გეგმა.

1.2. დიგიტალური ლექსიკოგრაფია

დიგიტალური ქართველოლოგიის განვითარება, როგორც ზემოთ უკვე აღვნიშნეთ, მოითხოვდა დიდი მონაცემების **ორი ტიპის ბაზების (big data)** შექმნას:

- **ტექსტური ბაზები** – აქ იგულისხმება როგორც სალიტერატურო ენის ტექსტური კორპუსები (მათ შორის თემატური კორპუსები, როგორიცაა, მაგალითად, იურიდიული კორპუსი, პოლიტიკური კორპუსი, წარწერების კორპუსი, სამეცნიერო მეტაენის კორპუსი და ა.შ.), ისე **ზეპირმეტყველების კორპუსები** (დიალექტების, სოციოლექტების და ა.შ. კორპუსები);
- **დიგიტალური ლექსიკონები** – როგორც ერთენოვანი (მაგ., განმარტებითი, დარგობლივი, ტერმინოლოგიური), ისე ორ- და მრავალენოვანი დიგიტალური ლექსიკონები, რაც თავის მხრივ **დიგიტალური ლექსიკოგრაფიის განვითარებას** მოითხოვდა.

ქრონოლოგიურად, დიგიტალური ლექსიკოგრაფიის განვითარებას ყოველთვის წინ უსწრებს **დიდ მონაცემთა ბაზების** შექმნა, ვინაიდან დიგიტალური ლექსიკონების დამუშავება არსებითად დიდ მონაცემთა ბაზების – ტექსტური კორპუსების ტექნოლოგიური „ტრანსფორმაციის“ შედეგად ხორციელდება. დიგიტალურ ჰუმანიტარიაში ტექსტური კორპუსების რესურსების ლექსიკოგრაფიულ რესურსად გარდაქმნის ორი ძირითადი მეთოდი არის აპრობირებული: ა) ტექსტიდან

ლექსიკონამდე ინდექსაციის გზით, და ბ) ტექსტიდან ლექსიკონამდე სიხშირულ ლექსიკონზე დაყრდნობით.

ლექსიკოგრაფიის დიგიტალიზაცია საქართველოში ორი სხვადასხვა მიმართულებით განვითარდა:

1. **რეტროდიგიტალიზაცია** – ლექსიკონების კომპიუტერიზაცია უკვე არსებული ბეჭდური ლექსიკონების გაციფრების გზით;
2. **დიგიტალური ლექსიკოგრაფიის განვითარება** – მონაცემთა დიდ ბაზებზე დაყრდნობით დიგიტალური ლექსიკონების გენერირება ლექსიკონის მუდმივი გადამუშავებისა და რედაქტირების შესაძლებლობით – ლექსიკონის მართვის პანელით.

რეტროდიგიტალიზაციის არსი იმაში მდგომარეობს, რომ რესურსების საყოველთაო წვდომის მიზნით ხორციელდება უკვე არსებული (ძირითადად ბეჭდური) რესურსების გაციფრება. რეტროდიგიტალიზაციის პრინციპი საფუძვლად დაედო ამერიკული კომპანიის Google LLC[93] მიერ ინიცირებული პროექტის Google books-ის კონცეფციას – პროექტის განხორციელების შედეგად შეიქმნა რესურსების რეპრეზენტაციული ბაზა, რომელიც მომხმარებელს ღია რესურსის სახით სთავაზობს ბეჭდური რესურსების უსარმაზარ კოლექციას: 2019 წელს, კომპანიის 15 წლის იუბილესთან დაკავშირებით, გამოქვეყნებული სტატისტიკის მიხედვით, კომპანიის მიერ შექმნილი ბაზის მოცულობა 40 მილიონზე მეტ ერთეულს მოიცავს, რომელიც დაახლოებით 400-მდე ენის რესურსს ასახავს.

რეტროდიგიტალიზაციის თვალსაჩინო მაგალითს წარმოადგენს ქართული ენის განმარტებითი ლექსიკონის რვატომეულის ელექტრონული ვერსია, რომელიც არნ. ჩიქობავას ინსტიტუტში შეიქმნა ენის მოდელირების ასოციაციასთან ერთად. ქეგლ-ის აღნიშნული ელექტრონული ვერსია რვატომეულის ელექტრონული ვერსიაა და ჩაენაცვლა ქეგლ-ის ბეჭდურ ვერსიას. შესაბამისად, ქეგლ-ის რესურსი უცვლელად გადავიდა ელექტრონულ ვერსიაში იმ სისტემური ხასიათის წინააღმდეგობებთან ერთად, რომელიც ბეჭდურ ვერსიაში დასტურდებოდა. მხედველობაში გვაქვს წინააღმდეგობრიობა რომელიც დასტურდება მასალის როგორც თეორიული, ისე ტექნიკური დამუშავების თვალსაზრისით (კერძოდ,

[93] ამერიკული ტექნოლოგიური კომპანია **Google Limited Liability Company**, რომელიც აქტიურად მოღვაწეობს ტექნიკისა და პროგრამული უსრუნველყოფის განვითარების სფეროში.

წინააღმდეგობით სალექსიკონო ერთეულის გრამატიკულ კვალიფიკაციასა და სიტყვა-სტატიაში მოყვანილ საილუსტრაციო მაგალითებს შორის) თვალსაჩინოებისათვის მოვიყვანთ რამდენიმე მაგალითს.

სალექსიკონო ერთეულად ზმნის შემთხვევაში ქეგლ-ში შეტანილია როგორც საწყისის ფორმა, ისე აწმყოს მესამე პირის ფორმა. ვნახოთ როგორი გრამატიკული კვალიფიკაცია ეძლევა საწყისის *„ბზუილი“* შესაბამის პირიან ფორმებს:

1. უზმნისწინო *ბზუის* ფორმა კვალიფიცირებულია, როგორც გარდაუვალი, **დინამიკურობა-სტატიკურობის** მიუთითებლად;
2. ზმნისწინიანი *დაბზუის* ფორმა კვალიფიცირებულია, როგორც გარდაუვალი **დინამიკური** ზმნა;
3. ზმნისწინიანი *მიბზუის* ფორმა კი კვალიფიცირებულია, როგორც გარდაუვალი **სტატიკური** ზმნა;

იგივე მდგომარეობა გვაქვს დინამიკურობა-სტატიკურობის თვალსაზრისით *„ხხაეილის“* შემთხვევაშიც:

შდრ.:



სალექსიკონო ერთეულის „ბზუის“ სიტყვა-სტატია ქეგლ-ში

შდრ.:



სალექსიკონო ერთეულის „დაბზუის“ სიტყვა-სტატია ქეგლ-ში

შდრ.:



სალექსიკონო ერთეულის „მიბზუის“ სიტყვა-სტატია ქეგლ-ში

შდრ.:



სალექსიკონო ერთეულის „ჩხავის“ სიტყვა-სტატია ქეგლ-ში (სტატიკური)

შდრ.:



სალექსიკონო ერთეულის „დასჩხავის“ სიტყვა-სტატია ქეგლ-ში (დინამიკური)

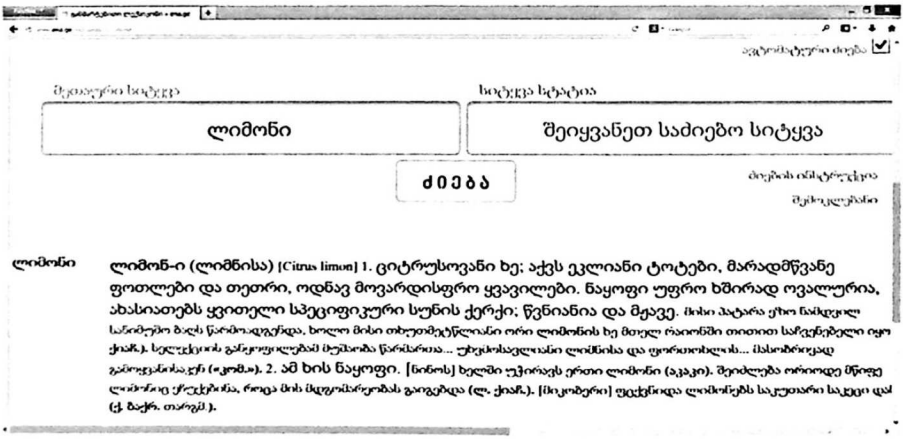
შდრ.:



სალექსიკონო ერთეულის „შესჩხავის“ სიტყვა-სტატია ქეგლ-ში (სტატიკური)

სისტემური თვალსაზრისით ძნელი ასახსნელია თუ რას ემყარება *დასჩხავის* და *მისჩხავის* ფორმების დიფერენციაცია დინამიკურობა-სტატიკურობის თვალსაზრისით. გრამატიკული კვალიფიკაციის თვალსაზრისით ზმნის აღნიშნულ პირიან ფორმებში ამგვარი განსხვავება გამოწვეულია იმ თეორიული ხასიათის დაპირისპირებით, რომელიც ცალკეული ტომის რედაქტორებს შორის არსებობდა.

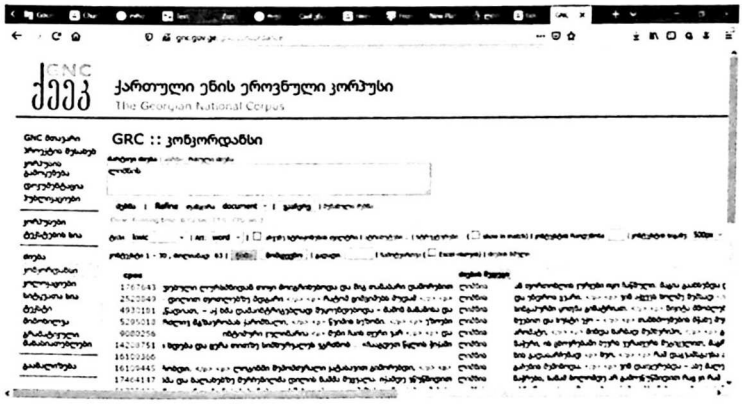
წინააღმდეგობრიობა და არათანმიმდევრულობა დასტურდება შიდასტრუქტურულ დონეზეც: სალექსიკონო ერთეულის გრამატიკულ კვალიფიკაციასა და სიტყვა-სტატიაში მოყვანილ საილუსტრაციო მაგალითებს შორის.



მაგალითად, სალექსიკონო ერთეული „ლიმონი“ განმარტებით ლექსიკონში დახასიათებულია როგორც ფუძეკუმშვადი: *ლიმონი (ლიმნისა)*, მაშინ როდესაც საილუსტრაციო მასალაში აღნიშნული სიტყვა ორჯერ დასტურდება უკუმშველი ფორმით (*ლიმონის ხე, ლიმონებს*), ერთხელ – კუმშვადი ფორმით (*ლიმნისა*).

ქართული ენის განმარტებით ლექსიკონში ორივე – კუმშვადი და უკუმშველი ფორმების არსებობა საილუსტრაციო მასალაში იმაზე მიუთითებს, რომ სალექსიკონო მასალების შეგროვების დროს სალიტერატურო ენაში ორივე ფორმა თანაარსებობდა. აღნიშნული ლექსიკური ერთეულის (*ლიმონი*) ბრუნების პარადიგმაში კუმშვადობის თვალსაზრისით ლაბილურობის ფაქტი (კუმშვადი და უკუმშველი ფორმების თანაარსებობა) ქართული ენის ეროვნულ კორპუსშიც დასტურდება – ***ლიმონის/ლიმნის*** სიტყვაფორმების სტატისტიკა ქეკ-ში ამგვარად გამოიყურება: უკუმშველი - ლიმონის (250) *vs.* ლიმნის (63)[94]. შდრ.:





[94] სტატისტიკური მონაცემის თარიღი: 02.01.2021

ქართული ენის განმარტებითი ლექსიკონის რვატომეულის ელექტრონული ვერსიის შექმნამ საფუძველი ჩაუყარა ბეჭდური და ელექტრონული ლექსიკონების დუბლეტს, ე.წ. **ჰიბრიდულ ლექსიკონებს**. ამგვარი ჰიბრიდული ორენოვანი ლექსიკონის მაგალითს წარმოადგენს **დონალდ რეიფილდის** რედაქტორობით გამოცემული **ქართულ-ინგლისური ლექსიკონის ორტომეული**, რომელიც 2006 წელს გამოქვეყნდა და 140 000 ლექსიკურ ერთეულს შეიცავს. ლექსიკონი კომპილაციური ხასიათისაა: ძირითადად იმეორებს რვატომეულის მასალას, თუმცა ლექსიკა განვრცობილია ქართული პრესის მონაცემთა ბაზაზე (1999-2002) დაყრდნობით, რამაც საგრძნობლად „გაათანამედროვა" ლექსიკონის ლექსიკური ბაზა. ლექსიკონში ასევე გამოყენებულია თანამედროვე მწერლების, ძირითადად ნოველისტების ტექსტები, ილია აბულაძისა და ზურაბ სარჯველაძის ძველი ქართული ენის ლექსიკონების მასალები და გამოყენებულია (შერჩევით) თანამედროვე ქართული ენის დიალექტების ოცდაათამდე ლექსიკონისა და სხვადასხვა სამეცნიერო-ტექნიკური დარგის ლექსიკონები. მანამდე გამოქვეყნებული ლექსიკონებისაგან განსხვავებით, აღნიშნული **ქართულ-ინგლისური ლექსიკონი** ასევე მოიცავს თანამედროვე სასაუბრო, ვულგარული გამოთქმების მნიშვნელოვან ნაწილს. მიუხედავად ამგვარი ლექსიკური მრავალფეროვნებისა, უცვლელი დარჩა ლექსიკონის აგების ტრადიციული, ჩიქობავასეული პრინციპი, რომელიც თავის დროზე საფუძვლად დაედო რვატომეულს, იმ განსხვავებით, რომ გრამატიკულ მახასიათებლებს აღნიშნულ ლექსიკონში დაემატა აქტანტიანობის რაოდენობრივი მაჩვენებელი: **Vt1** (გარდამავალი ზმნა ერთი არგუმენტით), **Vt2** (გარდამავალი ზმნა ორი არგუმენტით), **Vi2** (გარდაუვალი ზმნა ორი არგუმენტით) და ა.შ.

ჰიბრიდული ლექსიკონის ერთ-ერთ საინტერესო მაგალითს წარმოადგენს გიორგი ცოცანიძის, ნანა ლოლაძისა და ქეთევან დათუკიშვილის ავტორობით გამოცემული „ქართული ლექსიკონი", რომელიც ტექსტურ მონაცემთა ბაზების გადამუშავების საფუძველზე შეიქმნა და ელექტრონული ლექსიკონის სახით არსებულ ღია რესურსს წარმოადგენს[95].

ახალი ტიპის ლექსიკონის შექმნის ცდას წარმოადგენდა აკად. გ. წერეთლის სახ. „ვეფხისტყაოსნის" აკადემიური გამოცემისა და „ქართული ენის თესაურუსის" კომისიის მიერ განხორციელებული პროექტი „ქართული ენის თესაურუსი", რომელიც მიზნად ისახავდა საქართველოს მეცნიერებათა აკადემიის „ვეფხისტყაოსნის" აკადემიური ტექსტის დამდგენი კომისიის მიერ ათწლეულების განმავლობაში

[95] https://www.ganmarteba.ge/

დაგროვილი რესურსების გაციფრებას, კერძოდ, ქართული დამწერლობის ტექსტების ბარათებად გადაწერილი (დაახლოებით 6 მილიონი ბარათი) მასალისაგან ქართული ენის დოკუმენტირებული ლექსიკონის შექმნას. ლექსიკონი ნაწილობრივ ბეჭდური სახით არის გამოცემული[96]. პარალელურად მზადდება ლექსიკონის დიგიტალური ვერსია.

საქართველოში თანამედროვე დიგიტალური ლექსიკონების კონცეპტუალიზაცია თბილისის სახელმწიფო უნივერსიტეტის დასავლეთ ევროპის ენებისა და ლიტერატურის ფაკულტეტის ბაზაზე განხორციელდა და დაკავშირებულია პროფ. თინათინ მარგალიტაძის სახელთან. მისი ხელმძღვანელობით დამუშავდა ლექსიკონის გაწყობის როგორც სტრუქტურულ-ესენციალური პრინციპი, ისე ტექნოლოგიური ჩარჩო. სწორედ აღნიშნული კონცეფციაზე დაყრდნობით დაიწყო მუშაობა 1995 წელს ივ. ჯავახიშვილის სახელობის თბილისის სახელმწიფო უნივერსიტეტის ლექსიკოგრაფიულმა ცენტრმა დიდი ინგლისურ-ქართული ლექსიკონის (CEGD) სერიული გამოცემა. დღეისათვის გამოქვეყნებულია ინგლისურ-ქართული ლექსიკონის 14 ტომი, რომელიც მოიცავს 110 000 სალექსიკონო ერთეულის რამდენიმე ასეულ ათას ინგლისურ მნიშვნელობას, კოლოკაციებს, ფრაზულ ზმნებს, იდიომებს, ტერმინებს სხვადასხვა სფეროდან[97]. მოგვიანებით, აღნიშნული ბეჭდური გამოცემის ბაზაზე შეიქმნა ლექსიკონის ონლაინ-ვერსია, რომელიც 2010 წელს გამოქვეყნდა ინტერნეტში ღია რესურსის სახით[98].

ლექსიკონზე მომუშავე მეცნიერთა ჯგუფმა თავიდანვე უარყო ქართულ სამეცნიერო პრაქტიკაში არსებული, უცხოენოვანი სიტყვის მხოლოდ ერთი ეკვივალენტით წარმოდგენის პრინციპი ორენოვან ევროპულ-ქართული ტიპის ლექსიკონებში და ინგლისურ და ქართულ ენებზე სემანტიკური სისტემების აღწერისათვის შემოიღო ეკვივალენტურობის ორი დონე: მნიშვნელობის ეკვივალენტები და მთარგმნელობით-კონტექსტუალური ეკვივალენტები. მნიშვნელობის ეკვივალენტებისა და მთარგმნელობით-კონტექსტუალური ეკვივალენტების კომბინაციამ, განმარტებითი ლექსიკონების გარკვეული პრინციპების გამოყენებამ და საილუსტრაციო მასალის ზედმიწევნით ყურადღებით შერჩევამ საშუალება მისცა რედაქტორებს ლექსიკონში წარმოედგინათ სალექსიკონო ერთეულის გამოყენების

---

[96] http://science.org.ge/newsite/000/bookshelf/books/thesaurus.pdf?size=960x640

[97] იხ. თ. მარგალიტაძე, „ინგლისურ-ქართული ლექსიკოგრაფია“, სამეცნიერო ჟურნალი „ქართველოლოგი“, თბილისის სახელმწიფო უნივერსიტეტი, 2013.

[98] www.dict.ge

ყველა ასპექტი, ინგლისური სიტყვების მნიშვნელობის ყველა ნიუანსის გათვალისწინებით. ეს იყო ის სიახლე, რომელიც დიდი ინგლისურ-ქართული ონლაინლექსიკონის (CEGD)-ის შემქმნელმა გუნდმა შეიმუშავა და დანერგა ქართულ ლექსიკოგრაფიაში, რითაც საფუძველი ჩაუყარა დიგიტალური ლექსიკონების შექმნის თეორიულად და სისტემურად უაღრესად გამართულ ტრადიციას.

დიგიტალური ლექსიკოგრაფიის განვითარება საქართველოში არსებითად 2010 წლიდან იწყება და თანდათანობით იხვეწება სხვადასხვა ტიპის ლექსიკონების დამუშავების გზით, რომლებიც ქვეყნდება ელექტრონულ მატარებლებზე: როგორც ონლაინვერსია და როგორც CD-ROM-ი. ასეთი იყო, მაგალითად, დიდი ინგლისურ-ქართული ონლაინლექსიკონი (CEGOD), რომელიც ინტერნეტში გამოქვეყნდა 2010 წელს. მიუხედავად იმისა, რომ აღნიშნული ონლაინლექსიკონის მონაცემთა ბაზა DOC ფაილებისგან შეიქმნა, ტექნოლოგიური თვალსაზრისით იგი საკმაოდ მოქნილია: ლექსიკონი აღჭურვილია მომხმარებლის მხრიდან ლექსიკის მართვის ფუნქციებით, ლექსიკონში შესაძლებელია ორმხრივი ძიება, გვთავაზობს ავტომატურ შემოთავაზებებს და შესწორებებს. ლექსიკონის მართვის საგანგებო ფუნქცია საშუალებას აძლევს რედაქტორებს მოძებნონ და გაარედაქტირონ ლექსიკა, დაამატონ ლექსიკონში ახალი ჩანაწერები და ა.შ. აღნიშნული ლექსიკონის ტექნოლოგიური სისტემა ოდნავ შეცვლილი სახით გამოყენებული იქნა ლექსიკოგრაფიული ცენტრის შედარებით მცირე, სპეციალიზებული ლექსიკონებისთვის, როგორიცაა, მაგალითად, ინგლისურ-ქართული სამხედრო ონლაინლექსიკონი (http://mil.dict.ge), ინგლისურ-ქართული ბიოლოგიური ონლაინლექსიკონი" (http://bio.dict.ge). რაც შეეხება ლექსიკოგრაფიული ცენტრის მიერ გამოცემულ ინგლისურ-რუსულ-ქართული ტექნიკური ტერმინოლოგიის ონლაინლექსიკონს (2014 - 2016), იგი შექმნილია მრავალენოვანი ლექსიკონის მართვის სისტემაში[99], ისევე როგორც „ციფრული გზით დაბადებული" ორენოვანი ტერმინოლოგიური ონლაინ-ლექსიკონი[100].

---

[99] იხ.: თ. მარგალიტაძე, გ. ქერეჭაშვილი, „ეორდის ფაილებიდან თანამედროვე ონლაინლექსიკონამდე". საერთაშორისო კონფერენციის კრებული "ელექტრონული ლექსიკოგრაფია 21-ე საუკუნეში (Electronic lexicography in the 21st century: thinking outside the paper)". ტალინი, 2013. (http://eki.ee/elex2013); თ. მარგალიტაძე. „ქართული ტერმინოლოგიური ონლაინლექსიკონების ახალი პლატფორმა და მრავალენოვანი ლექსიკონის მართვის სისტემა". ევროპის ლექსიკოგრაფთა ასოციაციის (EURALEX) მე-18 საერთაშორისო კონგრესის კრებული. ლიუბლიანას სახელმწიფო უნივერსიტეტის გამომცემლობა. ლიუბლიანა, სლოვენია, 2018. (https://euralex.org/wp-content/themes/euralex)

[100] აღნიშნულ პროექტებს აფინანსებდა შოთა რუსთაველის ეროვნული სამეცნიერო ფონდი.

აღნიშნულმა მეთოდოლოგიურმა და ტექნოლოგიურმა წინსვლებმა თანამედროვე ქართულ ლექსიკოგრაფიაში ცალსახად გამოკვეთა დიგიტალური ლექსიკოგრაფიის ძირითადი პრინციპი: „ქართული მომავალი ციფრული ლექსიკონები უნდა შეიქმნას არა როგორც ბეჭდური ლექსიკონების ელექტრონული ვერსიები, არამედ როგორც კორპუსზე დაფუძნებული ლექსიკონები, რომლებიც სპეციალურად შექმნილია და მომზადებულია ლექსიკონის წერის სისტემებში (Dictionary Writing System, DWS) ელექტრონული და ონლაინფორმატისთვის. ეს საშუალებას მისცემს ქართველ ლექსოგრაფებს ლექსიკონის მასალებს დაამატონ უამრავი ფუნქციონალური საშუალება: სინონიმები, ანტონიმები, საილუსტრაციო წინადადებები, ნახატები, გამოიყენონ სხვადასხვა ფერები, შრიფტები, ლექსიკონებში განახორციელონ სხვა არალექსიკოგრაფიული მასალის ინტეგრაცია (მაგ., სავარჯიშოები) და ა.შ. ... მეტი ყურადღება უნდა დაეთმოს ლექსიკონის გამოყენების შესწავლას, რათა გათვალისწინებულ იქნას ლექსიკონის მომხმარებელთა საჭიროებები ახალი თაობის ქართული ლექსიკონების შემუშავებისა და დაგეგმვისას“[101].

დიდ მონაცემთა ბაზების შექმნისა და ლექსიკოგრაფიის დიგიტალიზაციის პროცესში საქართველოში ნელ-ნელა იკიდებს ფეხს კორპუსლინგვისტიკა. ქართულ ენაზე პირველად ჩნდება თეორიული ნაშრომი კორპუსლინგვისტიკაში,[102] მიმდინარეობს მუშაობა კორპუსლინგვისტიკის მეტაენის გასაქართულებლად,[103] ტარდება სამეცნიერო კონფერენციები და საზაფხულო სკოლები კორპუსლინგვისტიკაში[104], მრავალფეროვანი ხდება კორპუსების თემატიკაც. მომდევნო ქვეთავში დაწერილებით განვიხილავთ კორპუსლინგვისტიკის საკვანძო საკითხებს.

---

[101] იხ. T. Margalitadze, Lexicography of Georgian: a Brief Overview. Language@Leeds Working Papers in Linguistics, 1, University of Leeds. 2020, https://techdict.ge/

[102] იხ. მ. თანდაშვილი, ზ. ფურცხვანიძე, „ქართული ენის კორპუსლინგვისტური პარადიგმა“, ფრანკფურტი-თბილისი, გამომცემლობა ინტელექტი, 2014.

[103] მ. თანდაშვილი, ზ. ფურცხვანიძე, კორპუსლინგვისტიკის გლოსარიუმი, ფრანკფურტი-თბილისი, 2014.

[104] 2012 წელს ბათუმის უნივერსიტეტში მანანა თანდაშვილისა და რამაზ ხალვაშის ინიციატივით პირველად ჩატარდა საზაფხულო სკოლა კორპუსლინგვისტიკაში, რომელმაც საფუძველი ჩაუყარა ბათუმის საზაფხულო სკოლას დიგიტალურ ჰუმანიტარიაში. 2014-16 წლებში მანანა თანდაშვილისა და დარეჯან თვალთვაძის ინიციატივით სამჯერ ჩატარდა ზამთრის სკოლა დიგიტალურ ჰუმანიტარიაში. 2016 წელს ჩატარებული ზამთრის სკოლის ბაზაზე განხორციელდა სტუდენტური პროექტი „ქართული პოლიტიკური მეტყველების კორპუსი“, რომელმაც წინაპირობა შეუქმნა 2016-19 წლებში შექმნილ ქართული ენის ეროვნული კორპუსის თემატურ ქვეკორპუსს „პოლიტიკური სუპერმეტყველების კორპუსი“.

2. კორპუსლინგვისტიკის არსი, ამოცანები და მიზნები

ერთ-ერთი ცნობილი პრობლემა, რომელსაც ადგილი აქვს კორპუსლინგვისტიკასთან მიმართებაში, არის იმის განსაზღვრა, თუ რა არის კორპუსლინგვისტიკა: **დამოუკიდებელი დისციპლინა, კვლევის ახალი ინსტრუმენტი თუ მეთოდი?** ამის განსაზღვრა კი იმაზეა დამოკიდებული, თუ რას მივიჩნევთ კორპუსლინგვისტიკის კვლევის ობიექტად. ერთი შეხედვით, პასუხი მარტივია: კორპუსლინგვისტიკის კვლევის ობიექტი არის ენა. რით განსხვავდება მაშინ იგი ტრადიციული ლინგვისტიკისაგან? მხოლოდ იმით, რომ ენობრივი მონაცემების მანქანური დამუშავების ინსტრუმენტები გააჩნია და კვლევის ახალი, ტექნოლოგიური მეთოდებით არის აღჭურვილი?

ზოგადად, სამეცნიერო დისციპლინის არსებობაზე, როგორც წესი, მაშინ საუბრობენ, როდესაც ცალსახად არის გამოკვეთილი **კვლევის მიზანი, კვლევის საგანი და კვლევის საკუთარი მეთოდები.** ამიტომ, კორპუსლინგვისტიკის არსის დასაღგენად, უპირველეს ყოვლისა, სამ კითხვას უნდა გავცეთ პასუხი:

1. რა არის კორპუსლინგვისტიკის მიზანი?
2. რა არის კორპუსლინგვისტიკის კვლევის ობიექტი?
3. გააჩნია თუ არა კორპუსლინგვისტიკას კვლევის საკუთარი მეთოდები?

ჰუმანიტარულ მეცნიერებაში მონაცემთა თანამედროვე ტექნოლოგიური საშუალებებით დამუშავების აუცილებლობამ და მონაცემთა თანამედროვე ტექნოლოგიური საშუალებებით დამუშავების იმ შესაძლებლობებმა, რომელსაც ინფორმატიკა გვთავაზობს, ხელი შეუწყო ახალი დარგთაშორისი სამეცნიერო პროცესის – **კორპუსების შექმნისა და კორპუსული ლინგვისტიკის** განვითარებას.

ქართულ სამეცნიერო სივრცეში ტერმინი „კორპუსლინგვისტიკა“ თავდაპირველად სინტაგმის – **კორპუსის ლინგვისტიკა** – სახით შემოვიდა. ვინაიდან ჩვენ ეს შეთანხმება სემანტიკურ დონეზე პრობლემურად მიგვაჩნია (ლინგვისტიკა აქ დეტერმინირებულია როგორც კორპუსის შემსწავლელი მეცნიერული დისციპლინა), ტერმინის აღეკვატურობიდან გამომდინარე, ვამჯობინებთ კომპოზიტის **კორპუსლინგვისტიკა** გამოყენებას.

კორპუსლინგვისტიკის განსაზღვრება დღეს, გამომდინარე მისი კვლევის ობიექტიდან და მიზნებიდან, არაერთგვაროვანია. თეორიული შრომები კორპუსლინგვისტიკაში, როგორც წესი, პრაქტიკოსი კორპუსლინგვისტების მიერ არის დაწერილი, ამიტომ კორპუსლინგვისტიკის თეორიულ ლიტერატურაში არსებული განსაზღვრებების გვერდით მოვიტანთ კორპუსლინგვისტიკის განსაზღვრებებს იმ ინსტიტუციებისა

და სამეცნიერო პროექტების ვებგვერდებიდან, რომლებიც კორპუსების შექმნით არიან დაკავებული:

გერმანული რეფერენციული კორპუსი
(http://www1.ids-mannheim.de/kl.html)

„კორპუსლინგვისტიკის სამეცნიერო პროგრამა (მიზანი) მდგომარეობს იმაში, რომ დიდი მოცულობის ენობრივ მონაცემთა ბაზების ექსპლორატიული ანალიზის შედეგად მიღებული ცოდნის საფუძველზე ახალი თვალთახედვა შევიმუშაოთ ენის სტრუქტურის, წესების ფუნქციონირებისა და თვისებების შესახებ.“

ბერლინის ჰუმბოლდტის უნივერსიტეტის გერმანული ენისა და ენათმეცნიერების ინსტიტუტის ვებგვერდი
(https://www.linguistik.hu-berlin.de/institut/professuren/korpuslinguistik)

„კორპუსლინგვისტიკა შეისწავლის (ელექტრონული) კორპუსების აგების, შიდამოწყობის და მისი გამოყენების შესაძლებლობებს. კორპუსები წარმოადგენენ ტექსტების ნაკრებს, რომლებიც მიზანმიმართულად იქმნება იმ ლინგვისტური პრობლემატიკიდან გამომდინარე, რომლის კვლევასაც მიზნად ისახავს კონკრეტული კორპუსი. ჩვენ ვქმნით სპეციალურ კორპუსებს, როგორიცაა სასწავლო კორპუსი ან ისტორიული კორპუსი, და ამით ვქმნით კორპუსის არქიტექტურას.“

კომპიუტერული ლინგვისტიკის პორტალი (http://www.computerlinguistik.org/portal/portal.html?s=Korpuslinguistik)

„კორპუსლინგვისტიკა, უპირველეს ყოვლისა, არის მეთოდი, რომელიც საშუალებას გვაძლევს ლინგვისტური პრობლემატიკა დიგიტალურ ტექსტებზე დაყრდნობით ვიკვლიოთ. კვლევის განსახორციელებლად საჭიროა ენის წერილობითი და ზეპირმეტყველების ნიმუშების გაციფრება და მათი ანოტირება მეტყველების ნაწილების მიხედვით. კვლევის ზოგი ასპექტი მოითხოვს დიდი მოცულობის მონაცემების გადამუშავებას, რომელიც უფრო დეტალურ, ამასთან ავტომატურ ანოტაციას საჭიროებს.“

შტეფენ ბერმანის სალექციო კურსი კორპუსლინგვისტიკაში (ბოხუმის უნივერსიტეტი)
http://homepage.ruhr-uni-bochum.de/stephen.berman/Korpuslinguistik/Allgemeines.html)

„კორპუსლინგვისტიკა არის მეცნიერების ერთ-ერთი დარგი, რომელიც თავისი მიზნებიდან გამომდინარე, კვლევას ახორციელებს ავთენტურ ენობრივ მასალებზე დაყრდნობით – ანუ კვლევაში იყენებს უკვე არსებულ იმ მონაცემებს, რომლებიც არა-

ლინგვისტური მიზნებით არის შექმნილი, მაგ., ბულეტრისტიკა, პუბლიცისტიკა, დარგობრივი ლიტერატურა, პრესა, ზეპირი გამოსვლები, სპონტანური მეტყველება და ა.შ. კორპუსლინგვისტიკა არ იყენებს იმ მონაცემებს, რომლებსაც ლინგვისტები საკუთარი კომპეტენციის (ინტროსპექციის) საფუძველზე ქმნიან ან აგროვებენ ექსპერიმენტებისა და კითხვარების სახით კონკრეტული ენის მატარებლებისაგან, რათა ამ გზით შეამოწმონ საკუთარი თეორიული პოსტულატები.“

კორპუსლინგვისტიკის აქ მოტანილი რამდენიმე განსაზღვრებიდანაც კარგად ჩანს ის სირთულე, რომელიც თავს იჩენს კორპუსლინგვისტიკის განსაზღვრებასთან დაკავშირებით. კორპუსლინგვისტიკის, როგორც ცნების ცალსახა განმარტება, როგორც ასეთი, არ არსებობს. ამიტომ ამ საკითხს აქ აღარ ჩავუღრმავდებით, გვინდა მხოლოდ იმ რამდენიმე ასპექტზე გავამახვილოთ ყურადღება, რომელიც კამათს არ იწვევს:

1. კორპუსლინგვისტიკა ეყრდნობა ავთენტურ ენობრივ რესურსებს;
2. კორპუსლინგვისტიკა კვლევას ახორციელებს გაციფრებულ რესურსებზე დაყრდნობით;
3. კორპუსლინგვისტიკა იყენებს დიდი მოცულობის ემპირიულ მასალას;
4. კორპუსლინგვისტიკის კვლევის საგანია არა მარტო წერილობით დოკუმენტირებული, არამედ ზეპირმეტყველების ენა, ანუ ენის დაგეგმილი და დაუგეგმაელი ვარიანტები;
5. კორპუსების შექმნისა (კორპუსის არქიტექტურა) და მისი მოხმარების ტექნოლოგიზების გარდა კორპუსლინგვისტიკის ამოცანაა ენათმეცნიერებაში ახალი ლინგვისტური თვალთახედვის გახსნა და ინტეგრაციული კვლევის დანერგვა.

ზემოთქმულის გათვალისწინებით, კორპუსლინგვისტიკის მიზანი ასე შეიძლება ჩამოვაყალიბოთ: **კორპუსლინგვისტიკის მიზანია ავთენტური ენობრივი მონაცემების გაციფრება და მათი თავმოყრა მონაცემთა სტრუქტურირებულ ბაზებში (ანუ კორპუსების შექმნა), ენის შესახებ დაგროვილი ცოდნის განახლება-განვრცობა, არსებული თეორიულ-ლინგვისტური პოსტულატების ვერიფიცირება/გენერირება დიდი მოცულობის ტექსტური კორპუსების ბაზაზე, კვლევის შედეგად მიღებული ცოდნის ტექნოლოგიზება – პროგრამული პროდუქტების შექმნა და მათი დანერგვა როგორც თეორიულ, ისე პრაქტიკულ სამომხმარებლო სფეროში.**

ანუ, ტრადიციული ლინგვისტიკისაგან განსხვავებით კორპუსლინგვისტიკის ძირითად მიზანს წარმოადგენს:

- კორპუსის სტრუქტურირების ძირითადი პრინციპების დადგენა და მისი ეფექტიანი გამოყენების მექანიზმების შექმნა, ანუ კორპუსის ტექნოლოგიური ჩარჩოს გენერირება. ეს ტექნოლოგიური ჩარჩო კორპუსის თემატიკიდან და კვლევის ინტერესებიდან გამომდინარე იქმნება;
- ენობრივი ფაქტების დიგიტალიზაცია და მისი ვირტუალური თავშოყრა კორპუსში (ემპირიული **input**);
- ენობრივი ფაქტებისა და მოვლენების სისტემური კვლევა მისი გამოვლენის ნებისმიერ – წერილობით ან ზეპირმეტყველების – ფორმაში;
- უკვე არსებული ლინგვისტური ცოდნის სისტემური რევიზია და აქტუალიზაცია, როგორც თეორიული ლინგვისტიკის განვითარების ერთ-ერთი წინაპირობა (ე.წ. თეორიული **update**);
- კვლევის შედეგების "ტექნოლოგიზება" – კვლევის შედეგების ტექნოლოგიურ პროდუქტად გარდაქმნა, რომელსაც პრაქტიკული გამოყენება ექნება როგორც მეცნიერების სხვადასხვა დარგებში, ისე ფართო საყოფაცხოვრებო სამომხმარებლო სფეროში (ტექნოლოგიური **output**).

I	II	III	IV
Corpus building			
	Empirical input		
		Theoretical update	
			Technological output

ზემოთქმულიდან გამომდინარე, კორპუსლინგვისტიკის ამოცანა შეიძლება ოთხ ძირითად პუნქტად ჩამოვაყალიბოთ:

I. კორპუსის გენერირება (ტექნოლოგიური ჩარჩოს შექმნა)

II. კორპუსის შევსება რეპრეზენტაციულობის თვალსაზრისით (დაბალანსება)

III. ლინგვისტური თეორიების განახლება (აქტუალიზაცია)

IV. კორპუსის შედეგების ტექნოლოგიზება

კორპუსლინგვისტიკა საქართველოში ჩვენ თვალწინ ვითარდება. მისი განვითარების ერთ-ერთ მთავარ წინაპირობად სამი ასპექტი მიგვაჩნია:

1. თეორიული ბაზისის შექმნა – ქართული ენის ტექნოლოგიზების კონცეპტუალიზაცია;

2. **იურიდიული და ფინანსური უზრუნველყოფა** – სახელმწიფო მხარდაჭერა ქართული ენის ტექნოლოგიზების განსახორციელებლად;
3. **საგანმანათლებლო** – კორპუსლინგვისტიკის, როგორც სასწავლო კურსის უმაღლეს საგანმანათლებლო სისტემის პროგრამაში ინტეგრაცია და ახალგაზრდა თაობის მისანმიმართული აღზრდა.

მიუხედავად იმისა, რომ 2012 წლიდან მოყოლებული რამდენჯერმე იყო მცდელობა საქართველოს პარლამენტის მერვე მოწვევის თავმჯდომარის დავით უსუფაშვილისა და პარლამენტის წევრის ფრიდონ საყვარელიძის მხრიდან, სახელმწიფო დონეზე მომხდარიყო კორპუსლინგვისტიკასა და ენის ტექნოლოგიების სფეროში მომუშავე ჯგუფების გაერთიანება ტექნოლოგიური კონცეფციის შესამუშავებლად (ამ მიზნით საქართველოს პარლამენტის ბიბლიოთეკაში დაარსდა **ქართული ენისა და თანამედროვე ტექნოლოგიების ცენტრი**[105]), ამ საკითხის მოგვარება მაინც ვერ მოხერხდა ამ დარგში მომუშავე სპეციალისტთა თეორიულ და პიროვნულ დონეზე წარმოშობილი აზრთა სხვადასხვაობის გამო. კორპუსლინგვისტიკის კურსი რამდენიმე წლის განმავლობაში იკითხებოდა თბილისის სახელმწიფო უნივერსიტეტსა და რამდენიმე სხვა უნივერსიტეტში ფაკულტატიურ საგნად, მაგრამ მისი ინტეგრაცია სასწავლო პროგრამაში მაინც ვერ მოხერხდა იმ პერიოდში.

3. კორპუსლინგვისტიკის კვლევის ობიექტი

კორპუსლინგვისტიკა, ტრადიციული ლინგვისტიკის მსგავსად, ენობრივ ფაქტებს იკვლევს, ოღონდ მისგან განსხვავებით ა) დიდი მოცულობის ციფრულ მონაცემთა ბაზებს (**Big Date**) ეყრდნობა და ბ) კვლევის პროცესში თანამედროვე ტექნოლოგიებს იყენებს. ლინგვისტიკასა და კორპუსლინგვისტიკას შორის განსხვავება, ერთი შეხედვით, მხოლოდ მონაცემთა სიდიდესა და დამუშავების პროცედურის **ტექ**ნოლოგიზებაში (ანუ მონაცემთა ელექტრონული დამუშავების შესაძლებლობაში) მდგომარეობს; მაგრამ თუ საკითხს კორპუსლინგვისტიკის საგანთან მიმართებაში ჩავულრმავდებით, ადვილად აღმოვაჩენთ არსებით განსხვავებას ტრადიციულ ლინგვისტიკასა და კორპუსლინგვისტიკას შორის.

რა არის ტრადიციული ლინგვისტიკის კვლევის ობიექტი?

- ტრადიციული ლინგვისტიკა იკვლევს ენას მისი „სწორი" ფორმების (ანუ ნორმების) დადგენისა და შესაბამისად, ამ ენის გრამატიკის შექმნის მიზნით. მისი კვლევის რესურსს ძირითადად წარმოადგენს ნორმირებული ენით

[105] იხ. https://www.youtube.com/watch?v=vZB3tQKs9_o

გადმოცემული წერილობითი წყაროები.

რა არის კორპუსლინგვისტიკის კვლევის ობიექტი?

– კორპუსლინგვისტიკის კვლევის ობიექტი არის ენა, ოღონდ ენა არა როგორც ნიშანთა აბსტრაქტული სისტემა, არამედ როგორც ენობრივი ნიშნების გამოყენების შესაძლებლობების ჯამი.

კორპუსლინგვისტიკისათვის მნიშვნელოვანია არა მხოლოდ სალიტერატურო, ნორმირებული ენის ფაქტები, არამედ მისი გეოგრაფიული და სოციალური ინვარიანტების (დიალექტებისა და სოციოლექტების) ფართო რეგისტრიც; ანუ ის ენობრივი რესურსები, რომელსაც ეყრდნობა კორპუსლინგვისტიკა, შეიცავს **ენის ფუნქციონირების ყველა ბუნებრივ შემთხვევას**. შესაბამისად, კორპუსლინგვისტიკაში კვლევა ემყარება და ითვალისწინებს როგორც ნორმირებულ (წერილობით), ისე „არანორმირებულ" (ზეპირმეტყველების ნიმუშები) ენობრივ ფაქტებს. ანუ კორპუსლინგვისტიკა იკვლევს ენას მის ბუნებრივ მოცემულობაში – დინამიკაში.

აქედან გამომდინარე, კორპუსლინგვისტიკის საკვლევი ბაზა – ტექსტური კორპუსები – ისე უნდა იყოს აგებული, რომ ასახავდეს ენის რეალიზების რაც შეიძლება ფართო სპექტრს[106]. ანუ, ტექსტური კორპუსი არა მარტო რეპრეზენტაციულობით უნდა ხასიათდებოდეს (ანუ აკმაყოფილებდეს მოცულობით და თემატურ მოთხოვნილებებს), არამედ უნდა ასახავდეს ენის რეალიზაციის ჟანრულ მრავალფეროვნებასაც – ენის რეალიზების განსხვავებულ სახეებს.

ზემოთქმულიდან გამომდინარე, **კორპუსლინგვისტიკის მიზანია ენის რეალიზების ნებისმიერი ემპირიული ფაქტის კვლევა და ენათმეცნიერებაში გავრცელებული დიქოტომიის „Langue" – „Parole"-ის მოხსნა. შესაბამისად, მისი კვლევის ობიექტს წარმოადგენს არა მხოლოდ დაგეგმილი, არამედ დაუგეგმავი ენის ფაქტებიც[107].**

დასკვნის სახით შეგვიძლია ვთქვათ, რომ, არსებითი განსხვავება ლინგვისტიკასა და კორპუსლინგვისტიკას შორის ენის ფენომენოლოგიურ გაგებაში უნდა ვეძებოთ.

---

[106] იხ. Noah Bubenhofer, *Sprachgebrauchsmuster. Korpuslinguistik als Methode der Diskurs- und Kulturanalyse.* Berlin, New York: de Gruyter, 2009.

[107] განმარტება: დაგეგმილი ენა (**planned speech**) არის ენის რეალიზების ის ფორმა, რომელიც წერილობით დოკუმენტირებულ ტექსტებში ვლინდება და ასახავს ენის გარკვეულ რეესტრს – ნორმირებულ ენას. მისგან განსხვავებით, ზეპირმეტყველება ხასიათდება სპონტანურობით და ენობრივი ფორმების მოხმარების თავისუფლებით, რაც საშუალებას გვაძლევს, დავაკვირდეთ ენაში მიმდინარე ცვლილებებსა და ტენდენციებს ბუნებრივ გარემოში. ზეპირმეტყველების გზით რეალიზებულ სპონტანურ ენას დაუგეგმავი ენა (**unplanned speech**) ეწოდება.

4. რა არის კორპუსი?

კორპუსის შესახებ სამეცნიერო ლიტერატურაში არსებული განსაზღვრებები, ისევე როგორც კორპუსლინგვისტიკის განმარტება, ერთგვაროვანი არ არის. არაერთგვაროვნება კორპუსის ძირითადი რესურსების რაგვარობასთან არის დაკავშირებული. როგორც წესი, კორპუსი იგივდება ცნებასთან „ტექსტური კორპუსი". გარდა წმინდა ტექსტური კორპუსებისა, არსებობს ე.წ. „ენობრივი კორპუსები", რომელიც არა მხოლოდ ზეპირმეტყველების ტრანსკრიბირებულ, ტექსტურ ვერსიებს მოიცავს, არამედ შესაბამის აუდიო და ვიდეო მასალებსაც, აღჭურვილს ფონეტიკური და ლინგვისტური ანოტაციით. მაგალითისათვის შეგვიძლია დავასახელოთ გერმანულის ზეპირმეტყველების კორპუსი **Datenbank Gesprochenes Deutsch (DGD)**, რომელიც მანჰაიმის გერმანული ენის ინსტიტუტში (IDS) (http://dgd.ids-mannheim.de:8080/dgd/pragdb. dgd_extern.welcome) შეიქმნა და შედგება დაახლოებით 900 ვიდეოჩანაწერისგან და 16 300 აუდიოჩანაწერისგან. მანჰაიმის ზეპირმეტყველების მასალების ქრონომეტრაჟი მოლიანობაში შეადგენს 4400 საათს.

ჰ. ბუსმანის (H. Bussmann) „ენათმეცნიერების ლექსიკონის" მიხედვით „კორპუსი არის კონკრეტული ენობრივი გამოხატულებების სასღვრული რაოდენობა, რომელიც გამოიყენება როგორც ემპირიული საფუძველი ენათმეცნიერული კვლევებისათვის. კორპუსის მნიშვნელობა და აგებულება დამოკიდებულია სპეციფიური ამოცანებისა და თეორიული ჩარჩოს მეთოდურ წინაპირობებზე".[108] ამ განსაზღვრებაში ძალიან ზოგადად არის მოცემული კორპუსის არსი და მიზანი. ქვემოთ შევეცდებით, პრაქტიკოსი „კორპუსმეიკერების" გამოცდილების გათვალისწინებით, კორპუსის უფრო დეტალური განსაზღვრება ჩამოვაყალიბოთ.

კორპუსი (**corpus** – ლათ. *corpus*, მრ. *corpora* „სხეული") არის შემოკლებული აღნიშვნა ცნებისათვის *ტექსტკორპუსი* ან *ტექსტური კორპუსი.* კორპუსლინგვისტიკაში იგი აღნიშნავს რომელიმე ენის წერილობითი ძეგლების ან წერილობით დოკუმენტირებული ზეპირმეტყველების ნაკრებს, რომელიც, როგორც წესი, ელექტრონულ მატარებლებზე არის გადატანილი (ანუ გაციფრებულია) და მანქანურად დამუშავებადია – ტექსტები სტრუქტურირებულია საგანგებო ნიშნულებით. გარდა საკუთრივ პირველადი, ანუ ენობრივი მონაცემებისა, იგი შეიცავს მეორად ინფორმაციას – მეტამონაცემებს, აღჭურვილია ლინგვისტური ანოტაციით და გააჩნია კორპუსის მარ-

[108] Hadumod Bussmann, *Lexikon der Sprachwissenschaft*, 2008, Kröner, S. 143.

თვის სპეციალური სისტემა – კორპუსის მენეჯერი.[109]

კარმენ შერერი (Carmen Scherer) კორპუსის დეფინიციაში დამატებით ხაზს უსვამს იმ გარემოებას, რომ კორპუსი არის ავთენტური ენობრივი რესურსების ნაკრები.[110]

კორპუსი დიდი მოცულობით მონაცემთა დამუშავების სწრაფ და ეფექტურ საშუალებას წარმოადგენს. კორპუსების შექმნამ განსაკუთრებული მნიშვნელობა შეიძინა ისეთი ჰუმანიტარული დარგებისათვის, როგორიცაა ენათმეცნიერება, ლიტერატურათმცოდნეობა, ისტორიოგრაფია, თუმცა აქტიურად გამოიყენება სამართალმცოდნეობაში და სხვა ჰუმანიტარულ დარგებში.

კორპუსლინგვისტიკა იძლევა ცალკეული საკითხებისა და პრობლემების სისტემური კვლევის საშუალებას როგორც სინქრონიულ, ისე დიაქრონიულ ასპექტში, მაგრამ კორპუსლინგვისტიკა არ არის მხოლოდ „კორპუსოლოგია“, ანუ მეცნიერება **კორპუსთმშენებლობის** შესახებ. თუმცა, კორპუსის აგების მეთოდოლოგიის შექმნა – **კორპუსის სტრუქტურირების, აგების პრინციპებისა და მართვის მეთოდების შემუშავება** – კორპუსლინგვისტიკის ერთ-ერთ საკვანძო პრობლემად არის მიჩნეული.

კორპუსი გარკვეული შინაარსობრივი და სამეცნიერო კრიტერიუმების შედეგად შექმნილი გაციფრებული ტექსტების სტრუქტურირებული კრებულია, რომელიც აერთიანებს: ა) გარკვეული ჟანრისა და ტიპის ტექსტებს – (პირველად) მონაცემებს, და ბ) დამატებით მონაცემებს პირველადი მონაცემების შესახებ – მეორად მონაცემებს ანუ მეტამონაცემებს, გ) კორპუსის მართვის სპეციალურ სისტემას – კორპუსის მენეჯერს.

კორპუსის უმცირეს შინაარსობრივ ელემენტს წარმოადგენს **კომპონენტული ტექსტი**, რომელიც ელექტრონულ მატარებლებზე არის გადატანილი. ასეთი ტექსტი შეიძლება იყოს დიდი ან მცირე მოცულობის ლიტერატურული ძეგლის (მაგ., რომანი, ნოველა, პიესა, პოემა, ლექსი), სამეცნიერო ნაშრომის (მაგ., მონოგრაფია, სტატია, თეზისი), ნორმატიული დოკუმენტის (მაგ., კონსტიტუცია, ბრძანებულება), პრესის მასალის (მაგ., საგაზეთო სტატია) ან ზეპირი მეტყველების ნიმუშის გაციფრებული (დიგიტალური) ვერსია.

---

[109] Lothar Lemnitzer und Heike Zinsmeister, *Korpuslinguistik. Eine Einführung* (2., durchges. und aktual. Aufl.). Tübingen: Narr., 2010.

[110] Carmen Scherer, *Korpuslinguistik*. Heidelberg: Universitätsverlag Winter, 2006.

კორპუსი არ არის ტექსტური რესურსების უბრალო ნაკრები, ერთგვარი მათემატიკური ჯამი. ტექსტის სახით მოცემული ნებისმიერი ენობრივი რესურსი წინასწარ განიცდის დამუშავებას, სანამ კორპუსის რესურსად იქცევა.

კორპუსის შექმნის პროცესი ეტაპობრივად მიმდინარეობს:

- უპირველეს ყოვლისა, იქმნება ტექსტის ელექტრონული ვერსია კოდირების საერთაშორისო სტანდარტების (UNICODE) შესაბამისად, ანუ ადგილი აქვს ტექსტის ციფრულ ფორმატში გადაყვანას საერთაშორისო სტანდარტის შესაბამისად; ელექტრონული ტექსტი, როგორც წესი, კონვერტირდება HTML, შემდეგ კი XML ფორმატში.
- ელექტრონული ვერსია ზუსტად უნდა ასახავდეს დედნის სტრუქტურას. ამიტომ, ტექსტის დიგიტალური ვერსიის შექმნის შემდეგ აუცილებელია ტექსტის შიდასტრუქტურული დამუშავება რეფერენციის თვალსაზრისით – ელექტრონული ვერსიის დამუშავება ტექსტის მარკირების საგანგებო ნიშნულებით (textual markup).
- ტექსტის გარე რეფერენციის მიზნით ხდება ტექსტის მეტამონაცემების თვალსაზრისით დამუშავება – ტექსტთან მიმართებაში რელევანტური მეტამონაცემების შეტანა საგანგებო ბაზაში, რათა მომხმარებელს კორპუსის მასალების რელევანტური ნიშნების მიხედვით მოძებნა გაუადვილდეს.

მხოლოდ ამგვარად დამუშავებული ტექსტი შეიძლება მივიჩნიოთ კორპუსის უმცირეს შინაარსობრივ ელემენტად.

ტექსტური კორპუსების სახით საქმე გვაქვს ერთგვარ „საცავთან“, რომელიც მასში ინტეგრირებული კორპუსული ელემენტების გარკვეული ნიშნის მიხედვით მოძებნის საშუალებას იძლევა. კვლევის პროცესში მისი ეფექტურად გამოყენება მხოლოდ მაშინ არის შესაძლებელი, თუ მას თან ახლავს დამატებითი ინფორმაცია – მეტამონაცემები ტექსტის ენობრივი კუთვნილების, წარმოშობის (ორიგინალური, ნათარგმნი), ავტორის, დათარიღების და სხვა რელევანტური ნიშნების მიხედვით.

კორპუსთან მიმართებაში ხშირად ისმის კითხვა: რა არის კორპუსი, ელექტრონული ბიბლიოთეკა?

ამ კითხვაზე ცალსახად უნდა გავცეთ პასუხი: არა! კორპუსი არ არის ბიბლიოთეკის (ანუ ტექსტების ნაკრების) ელექტრონული ვარიანტი. კორპუსში ჩვენ ეძებთ არა ავტორებს ან ტექსტებს, არამედ ენობრივ ფაქტებს და ვიკვლევთ მათ შორის არსებულ ურთიერთმიმართებებსა და განპირობებულობებს.

კორპუსის გამოყენებას უზრუნველყოფს კორპუსის მართვის საგანგებო სისტემა – **კორპუსის მენეჯერი**. კორპუსის გამოყენების ეფექტიანობა დამოკიდებულია კორპუსმენეჯერის სრულფასოვან ფუნქციონირებაზე. კორპუსის აქ დასახელებულ სამიეე კომპონენტს (**პირველადი მონაცემი, მეტამონაცემი და კორპუსმენეჯერი**) დაწერილებით განვიხილავთ მომდევნო ქვეთავებში.

5. კორპუსის მახასიათებლები

კორპუსების შექმნის პროცესში განსაკუთრებული ყურადღება ექცევა კორპუსის არსებით მახასიათებლებს: კორპუსის მოცულობას, კორპუსის სახეს, კორპუსის ტიპს, კორპუსის დაბალანსებას, კორპუსის ვიზუალიზაციას. ქვემოთ განვიხილავთ ამ მახასიათებლებს ცალ-ცალკე.

5.1. კორპუსების მოცულობა

კორპუსის ერთ-ერთ არსებით მახასიათებელ ნიშანს წარმოადგენს მოცულობა, რომელიც **ტოკენებში** იზომება. ტოკენი (**token**) კორპუსის უმცირესი კონსტიტუენტური ელემენტია და ტრადიციულ ლინგვისტიკაში მას შეესაბამება ცნება „სიტყვაფორმა“. იდენტური სიტყვაფორმები ქმნიან **ტიპებს**. ტიპი (**type**) არის კორპუსლინგვისტიკის ერთეული, რომელიც სტატისტიკური მიზნით გამოიყენება და როგორც წესი, უდრის **n**-ტოკენს. ტოკენებში აღიწერება კორპუსების მოცულობა კვანტიტატიური თვალსაზრისით, ტიპების შემთხვევაში კი ხდება იდენტური ელემენტების (სიტყვაფორმების) დაჯგუფება. კორპუსლინგვისტიკაში ტიპიზაცია გულისხმობს იდენტური ელემენტების ჯგუფების აღწერას კვანტიტატიური თვალსაზრისით.

მაგალითისათვის განვიხილოთ გალაკტიონ ტაბიძის ცნობილი ლექსის ერთი სტროფი:

„ქარი ჰქრის, ქარი ჰქრის, ქარი ჰქრის,
ფოთლები მიჰქრიან ქარდაქარ,
ხეთა რიგს, ხეთა ჯარს რკალად შლის,
სადა ხარ, სადა ხარ, სადა ხარ.”

პირველი სტრიქონი შედგება 6 ტოკენისაგან, მაგრამ 2 ტიპისაგან – **ქარი** და **ჰქრის**. მეორე სტრიქონში სამი ტოკენი და სამი ტიპია (**ფოთლები, მიჰქრიან** და **ქარდაქარ**), მესამე სტრიქონში 6 ტოკენი და 5 ტიპია (**ხეთა, რიგს, ჯარს, რკალად** და **შლის**), ხოლო მეოთხე სტრიქონში – 6 ტოკენი და 2 ტიპი (**სადა** და **ხარ**).

ტოკენი და ტიპი საშუალებას გვაძლევს არა მარტო კორპუსის მოცულობა განვსაზღვროთ, არამედ დავალაგოთ იგი **ტიპების სიხშირის ზრდადობის** ან **კლებადობის** მიხედვით. ტოკენისა და ტიპის სტატისტიკა, როგორც კორპუსის რაოდენობრივი მახასიათებელი, არსებით მნიშვნელობას იძენს კორპუსის დაბალანსების დროს.

შედარებისათვის იხილეთ რამდენიმე კორპუსის მოცულობა[111]:

გერმანულის რეფერენციული კორპუსი – მოცულობა: 24 მილიარდი ტოკენი
Deutsches Referenzkorpus DeReKo des Instituts für Deutsche Sprache (IDS)
http://www.ids-mannheim.de/kt/projekte/korpora/

მე-20 საუკუნის გერმანულის რეფერენციული კორპუსი – მოცულობა: 100 მილიონი ტოკენი
Referenzkorpus der deutschen Sprache des 20. Jahrhunderts (DWDS-Kernkorpus) http://www.dwds.de/

კორპუსი C4[112] – მოცულობა: 46 მილიონამდე ტოკენი
Corpus C4
http://www.korpus-c4.org

ვებ-კორპუსი – მოცულობა: 9 მილიარდზე მეტი ტოკენი
COW: Corpora from the Web
http://hpsg.fu-berlin.de/cow/

გერმანული ტექსტების არქივი – მოცულობა: 100 მილიონამდე ტოკენი (გერმანული ტექსტების ისტორიული კორპუსი, 1600-1900 წლები, 1300 წიგნი)
Deutsches Textarchiv
http://www.deutschestextarchiv.de

ბრიტანული ეროვნული კორპუსი – მოცულობა: 100 მილიონი ტოკენი
British National Corpus (BNC)
http://www.natcorp.ox.ac.uk/

---

[111] მონაცემები აღებულია ნოა ბუბენჰოფერის (Noah Bubenhofer) ინტერნეტ-პორტალიდან "კორპუსლინგვისტიკის შესავალი".
იხ. http://www.bubenhofer.com/korpuslinguistik/kurs/index.php?id=korpustypen.html

[112] ერთობლივი კორპუსი, რომელიც მოიცავს მე-20 საუკუნის გერმანული ენის დიგიტალურ ლექსიკონს (DWDS), ავსტრიის აკადემიის კორპუსს (AAC), სამხრეთ ტიროლის კორპუსს და შვეიცარიული ტექსტების კორპუსს (CHTK).

თანამედროვე ამერიკული ინგლისურის კორპუსი – მოცულობა: 450 მილიონი ტოკენი
The Corpus of Contemporary American English (COCA)
http://corpus.byu.edu/coca/

ტექსტი+მთის კორპუსი (ალპინისტური ლიტერატურის კორპუსი) – მოცულობა: 37 მილიონამდე ტოკენი
Text+Berg-Korpus: Alpinistische Literatur (Jahrbuch des Schweizer Alpenclubs, Echo des Alpes, "Alpen")
http://www.textberg.ch

საზოგადოებრივ-სოციალური მოძრაობების კორპუსი – მოცულობა: 1 მლნ. ტოკენი
CosMov (Corpora for Social Movement Research)
http://www.semtracks.org/cosmov/

დორტმუნდის ჩატის კორპუსი – მოცულობა: 1,1 მილიონი ტოკენი, 150 000 ჩატ-გზაენილიდან
Dortmunder Chatkorpus
http://www.chatkorpus.uni-dortmund.de/

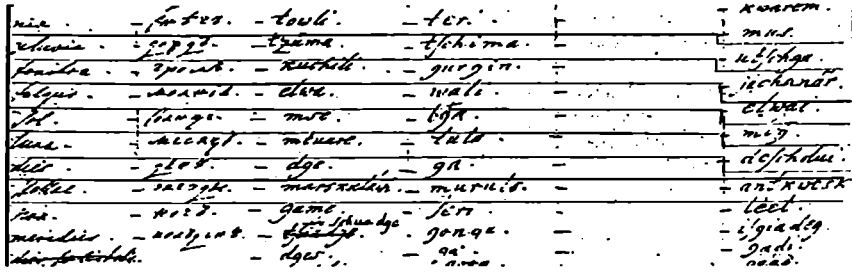
ამერიკული სერიალების (საპნის ოპერების) კორპუსი – მოცულობა: 100 მილიონი ტოკენი
Corpus of American Soap Operas
http://corpus2.byu.edu/soap/

5.2. კორპუსის სახეები

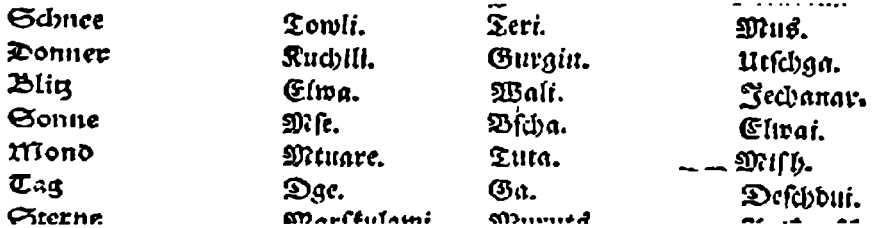
კორპუსის ცნება ჰუმანიტარულ დარგებში *a priori* ასოცირდება დიგიტალურ მონაცემთა ბაზებთან. თანამედროვე კორპუსები ელექტრონული რესურსებისაგან შედგება და აქედან გამომდინარე, იმპლიკაციურად მოიცავენ მსაზღვრელს „ელექტრონული". შესაბამისად, კორპუსის ცნება დღეისათვის უტოლდება ცნებას „ელექტრონული კორპუსი".

მიუხედავად იმისა, რომ **კორპუსის** ცნება იმ გაგებით, რა გაგებითაც იგი კორპუსლინგვისტიკაში გამოიყენება, მე-20 საუკუნის სამეცნიერო ტერმინი გახლავთ, შესაძლებლად მიგვაჩნია მისი მეტაფორული გამოყენება ქაღალდზე დოკუმენტირებული ტექსტური რესურსების აღსანიშნავადაც. თუ კორპუსი დღევანდელი გაგებით ელექტრონულად დოკუმენტირებულ რესურსთა ნაკრებს წარმოადგენს, მაშინ ქაღალდზე დოკუმენტირებულ რესურსთა მიზნობრივ ნაკრებს შეგვიძლია მეტაფორულად „ქაღალდის კორპუსი" ვუწოდოთ.

კავკასიური ენების დოკუმენტაციის ადრინდელი საფეხური შეიძლება განვიხილოთ როგორც ქაღალდის კორპუსის ეპოქა. XVII, XIX და XX საუკუნის პირველ ნახევარში ენობრივი მასალების დოკუმენტაცია, მაგნიტური მატარებლების არარსებობის გამო, არსებითად პირდაპირ ქაღალდზე ხდებოდა. ასე შეიქმნა ანტონ გიულდენშტედტის, იულიუს პალასის, ეელია ჩელების და სხვათა მიერ დოკუმენტირებული ენობრივი რესურსები.



ა. გიულდენშტედტის სიტყვანი ქართველური ენებისათვის, ხელნაწერი



ა. გიულდენშტედტის სიტყვანი ქართველური ენებისათვის, ბეჭდური ვერსია

ქაღალდის კორპუსის მაგალითად შეიძლება ჩაითვალოს მე-20 საუკუნის 30-იან წლებში მოპოვებული დიალექტური მასალების უნიკალური კოლექცია, რომელიც ივ. ჯავახიშვილის თაოსნობით მოწყობილი ისტორიული ექსპედიციის მიერ იქნა შეგროვებული და მიზნად ისახავდა საქართველოს ყველა კუთხის მეტყველების დოკუმენტაციას და დღეისათვის, სამწუხაროდ, მხოლოდ საარქივო მასალას წარმოადგენს.

ელექტრონული კორპუსების ერა გასული საუკუნის 60-იანი წლებიდან იწყება. ნელსონ ფრენსისისა (Nelson Francis) და ჰენრი კუჩერას (Henry Kučera) ავტორობით 1967 წელს გამოქვეყნებული სტატია, რომელმაც საბოლოოდ შეცვალა ლინგვისტური პარადიგმა – *Computational Analysis of Present-Day American English* – ეყრდნობოდა მათ მიერვე ინიცირებულ და განხორციელებულ ამერიკული ინგლისური ენის მონაცემთა ელექტრონულ ბაზას. ამ ელექტრონულ მონაცემთა ბაზის პირველი ვერსია 1960 წელს შეიქმნა და იმის გამო, რომ ფრენსისმა და კუჩერამ 1963 წლის თებერვლის

კონფერენციაზე Brown-ის უნივერსიტეტში საჯარო დისკუსიის საგნად აქციეს კორპუსის საბოლოო დიზაინის პრინციპები, ცნობილია სახელით **ბრაუნის კორპუსი**. აქვე შეჯერდა ექსპლიკატების საბოლოო ეტიკეტები და ისიც, რომ ტექსტები არა მარტო ჟანრულად, არამედ ქრონოლოგიურადაც შეზღუდულიყო. ის ფაქტი, რომ **ბრაუნის კორპუსი** ერთი წლის მანძილზე (1961) ამერიკაში გამოცემულ ტექსტურ პროდუქციას აერთიანებას, სწორედ ამ გარემოებით არის განპირობებული.

ბრაუნის კორპუსი, მიუხედავად მისი – დღევანდელი გადასახედიდან – მცირე მოცულობისა (1 მლნ. ტოკენი), უკვე შექმნის დროიდანვე ახერხებდა ყველა ფუნდამენტური კორპუსული ფუნქციის შესრულებას, სიტყვათა სიხშირის ასახვიდან დაწყებული ჰიპოთეზების სტატისტიკური შემოწმებით დამთავრებული.

თუ გავითვალისწინებთ იმ გარემოებას, რომ **ბრაუნის კორპუსის** შექმნის ფაზების უდიდესი ნაწილი არა ავტომატურ რეჟიმში, არამედ ხელით ხორციელდებოდა, ნათელი გახდება შესრულებული სამუშაოს კოლოსალური მასშტაბი. ენობრივი მონაცემების ასახვის ბალანსმა **ბრაუნის კორპუსში**, და ამ მონაცემების შემოფარგლამ ამერიკული ინგლისურის მოხმარების ერთწლიან ფარგლებში, პირველად დააფუძნა კორპუსლინგვისტური წარმოდგენა ე. წ. „ოქროს სტანდარტის" შესახებ, რომლისგანაც **ბრაუნის კორპუსი** მართალია პრაქტიკულად ჯერ კიდევ შორს იყო, თუმცა ცალსახად ასახა ამგვარი სტანდარტის მიღწევის გზა კორპუსლინგვისტიკაში.

კორპუსის სტრუქტურისა და კორპუსის გამოყენების მიზნებიდან გამომდინარე, განასხვავებენ კორპუსის სხვადასხვა სახეს:

ა) რეფერენციული კორპუსი
ბ) სტატიკური კორპუსი *vs.* მონიტორული კორპუსი
გ) ანოტირებული კორპუსი
დ) ერთენოვანი კორპუსი *vs.* მრავალენოვანი პარალელური კორპუსი
ე) სპეციალური ანუ თემატური კორპუსი.

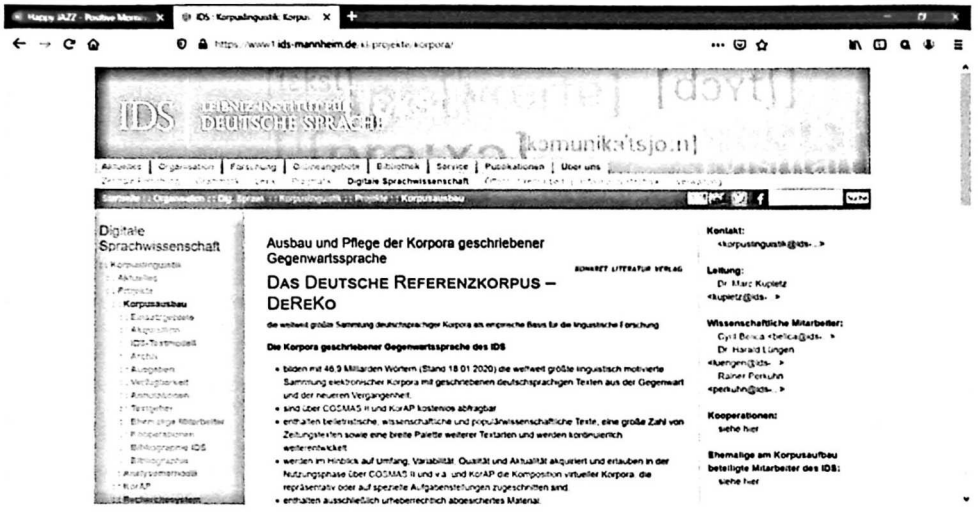
ქვემოთ სათითაოდ განვიხილავთ კორპუსის ამ სახეებს.

ა) რეფერენციული კორპუსი

რეფერენციული კორპუსი არის კორპუსის სახე, რომელიც მასში თავმოყრილი მასალების რეფერენციულობის მაღალი ხარისხით ხასიათდება – იგი, როგორც წესი, მიემართება სხვადასხვა კორპუსებს. რეფერენციული კორპუსის მაგალითად შეიძლება განვიხილოთ გერმანული ენის რეფერენციული კორპუსი **DeReKo**.

DeReKo (Das Deutsche Referenzkorpus)-ის შექმნა დაიწყო გასული საუკუნის 60-იანი წლებში მანჰაიმის გერმანული ენის ინსტიტუტში და დღესდღეობით გერმა-

ნული ენისათვის არსებულ ყველაზე რეპრეზენტაციულ კორპუსს წარმოადგენს. მისი მოცულობა 1992 წლიდან (28 მილიონი ტოკენი) 2020 წლამდე საგრძნობლად გაიზარდა (46,9 მილიარდი ტოკენი).



გერმანულ რეფერენციულ კორპუსში განთავსებული რესურსები საკმაოდ მრავალფეროვანია – გარდა მე-20 და 21-ე საუკუნეების ბელეტრისტიკისა, იგი მოიცავს სამეცნიერო, სამეცნიერო-პოპულარულ, დარგობრივ ლიტერატურას, ასევე პრესის მასალებს და სხვადასხვა თემატურ და ფრაგმენტულ კორპუსებს, როგორიცაა ბონის საგაზეთო კორპუსი (Bonner Zeitungskorpus), გოეთეს კორპუსი (Goete-korpus), გრამატიკის კორპუსი (Grammatik-korpus), გრიმის კორპუსი (Grimm-korpus), მარქსისა და ენგელსის კორპუსი (Marx-Engels-korpora), პოლიტიკური გამოსვლების გერმანული კორპუსი ვიზუალიზაციის თანხლებით, საპარლამენტო პლენარული სხდომების პროტოკოლების კორპუსი (Plenarprotokolle), ე.წ. „პოლიტიკური შემობრუნების“ კორპუსი (Wendekorpus), ინტერვიუების კორპუსი, ისტორიული კორპუსი და სხვა.

ბ) სტატიკური კორპუსი

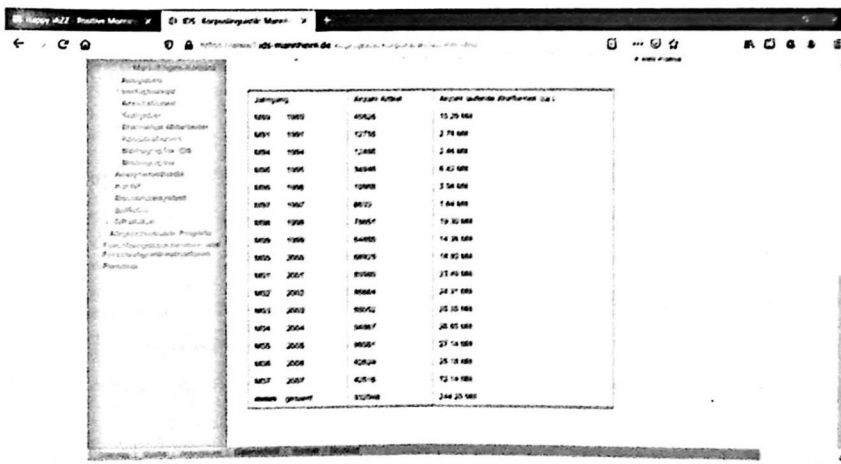
კორპუსების შექმნა მეცნიერთა და ტექნიკური პერსონალის მხრიდან მრავალწლიან მუშაობას მოითხოვს. კორპუსის შექმნის პირველსავე ეტაპზე განისაზღვრება კორპუსის ტიპი და დგინდება კორპუსში შემავალი მასალების მოცულობა. მაგ., თუ საქართველოს ტერიტორიაზე დაცული ქართული ეპიგრაფიკული წარწერების კორპუსის შექმნას გადაეწყვეტა, მაშინ კორპუსის რესურსად მოიაზრება როგორც ყველა ეპიგრაფიკული წარწერა, რომელიც უკვე დოკუმენტირებულია და ასა-

ხულია სამეცნიერო ლიტერატურაში, ასევე მუზეუმებში დაცული, მაგრამ ჯერ არგამოქვეყნებული წარწერების ჯგუფი, ისევე როგორც საქართველოს მთელ ტერიტორიაზე თავისუფლად გაბნეული და ჯერ არდოკუმენტირებული წარწერები. ამგვარი კორპუსის შექმნის შემთხვევაში საქმე გვაქვს სტატიკურ კორპუსთან, ანუ კორპუსთან, რომელიც თვისობრივად და რაოდენობრივად (მასში შესული მასალების თვალსაზრისით) აღარ შეიცვლება. ასეთ შემთხვევაში კორპუსი სტატიკურ მდგომარეობაში გადადის, ანუ აღარ მიმდინარეობს მისი თვისობრივი „განვითარება“ ან მასალობრივი განვრცობა. და რაც მთავარია, უცვლელი რჩება კორპუსის მენეჯერი – კორპუსის შიდა სტრუქტურა და საძიებო სისტემა.

სტატიკური კორპუსის არსებობა შეიძლება განპირობებული იყოს ორი ფაქტორით: კორპუსის რესურსების სასრულობით, თუმცა ეს ფაქტორი იმდენად არ არის გადამწყვეტი, რამდენადაც კორპუსის განვითარების პერსპექტივის უქონლობა. მაგ., შეიძლება ეპიგრაფიკული წარწერების კორპუსში მასალების სისრულეს მივაღწიოთ და მაქსიმალურად ამოვწუროთ რესურსების ასახვის შესაძლებლობა კორპუსში, მაგრამ კორპუსის განვითარება ამით არ დამთავრდეს – ტექნოლოგიური საშუალებების განვითარების შესაბამისად მოხდეს კორპუსის მუდმივი ტექნოლოგიური განახლება (update).

გ) მონიტორული კორპუსი

მონიტორული კორპუსი არის კორპუსის ისეთი სახე, რომელშიც მონაცემთა შეტანა პერიოდულად ახლდება, იცვლება სტრუქტურა და მოხმარების ინტერფეისი. მონიტორული კორპუსის მაგალითად გამოდგება მანჰაიმის საგაზეთო კორპუსი **Mannheimer Morgen.**



კორპუსის ის ნაწილი, რომელიც 1997 წლამდე გამოცემულ მასალებს ასახავს, ხშირად ფრაგმენტულია და არათანმიმდევრული, რაც არაელექტრონული გამოცემების პერიოდს ემთხვევა. მანჩაიმის საგაზეთო კორპუსი მონიტორული კორპუსია, ვინაიდან მუდმივად მიმდინარეობს მისი მონიტორინგი და კორპუსის ნაკლული ადგილების შევსება.

მანჩაიმის საგაზეთო კორპუსის მონიტორინგი თვალსაჩინოდაა ნაჩვენები ზემოთ მოცემულ სურათზე.

დ) პარალელური კორპუსი

კორპუსების უმრავლესობას მონოლინგვური კორპუსები შეადგენს. ამგვარი კორპუსების რესურსები მხოლოდ ერთი ენის მონაცემებს შეიცავს. თუმცა კომპარატივისტიკის დარგის მოთხოვნილებისა და შედარებითი ენათმეცნიერებისა თუ ლიტერატურათმცოდნეობის გაღრმავების, ასევე მთარგმნელობითი კვლევის ასპექტებიდან გამომდინარე, სულ უფრო გახშირდა ე.წ. პარალელური კორპუსების შექმნის პრეცედენტი.

პარალელური კორპუსი შეიძლება იყოს **მონოლინგვური** (ერთენოვანი პარალელური კორპუსი), **ბილინგვური** (ორენოვანი პარალელური კორპუსი) ან **მულტილინგვური** (მრავალენოვანი პარალელური კორპუსი).

მონოლინგვური პარალელური კორპუსების შექმნა განპირობებულია ისტორიული ძეგლების ლინგვისტური და კომპარატივისტული, ისევე როგორც ტექსტოლოგიური კვლევის სპეციფიკიდან გამომდინარე. საბეჭდი საშუალებების შექმნამდე და ქართული ნაბეჭდი წიგნების გამოცემამდე ქართული საეკლესიო და საერო ლიტერატურა ხელნაწერი წიგნების სახით ვრცელდებოდა, რომლებშიც შესაბამისად აისახა გადაწერის დროისათვის ენაში მიმდინარე ცვლილებები. შესაბამისად, ჩვენამდე მოღწეული სხვადასხვა საუკუნეში გადაწერილი ხელნაწერები ერთმანეთისაგან განსხვავდებიან ენობრივი (ფონოლოგიური, მორფოლოგიური, სინტაქსური, ლექსიკური) თვალსაზრისით. იმის საკვლევად, თუ რომელი ხელნაწერი უნდა მივიჩნიოთ საბაზისო ხელნაწერად და რომელი მათგანი გადაწერილად, ხოლო დაუთარიღებელი ხელნაწერების შემთხვევაში სისტემურად ვიკვლიოთ, რომელი ხელნაწერია უფრო ძველი, ფასდაუდებელია პარალელური კორპუსების არსებობა.

მონოლინგვური პარალელური კორპუსი იქმნება ერთი და იმავე ტექსტის ერთ ენაზე არსებული სხვადასხვა ვერსიების შეპირისპირებისა და დაპარალელების საფუძველზე. ერთენოვანი კორპუსები ძირითადად ისტორიულ ან ლიტერატურულ

ძეგლთა ხელნაწერების ვარიანტებისაგან იქმნება. ამგვარი კორპუსები შეიძლება შეიქმნას ლიტერატურული, რელიგიური ან ისტორიული ძეგლებისათვის, როგორიცაა, მაგალითად, **ოთხთავის ქართული თარგმანების პარალელური კორპუსი Tetraevangelium (*synopsis completa*)**[113], რომელიც ქართული ხელნაწერების დღეისათვის არსებულ ერთ-ერთ რეპრეზენტაციულ დიგიტალურ ბაზას წარმოადგენს.

დღეისათვის არაერთი პარალელური კორპუსი არის შექმნილი ქართველოლოგიური კვლევების განსახორციელებლად. ქვემოთ დავასახელებთ რამდენიმე მათგანს:

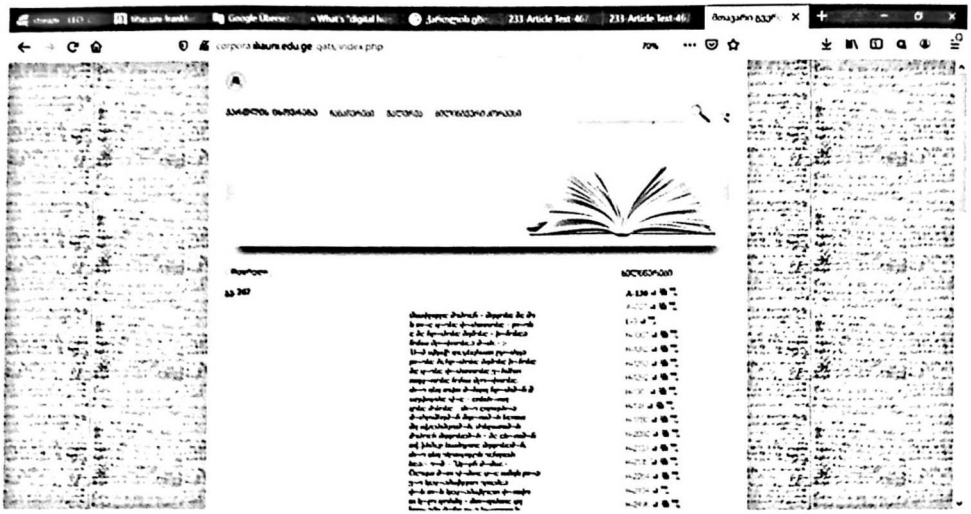
– **ქართულ-ინგლისური პარალელური კორპუსი – არიანე ჭანტურიას თარგმანები** (თსუს ლექსიკოგრაფიის ცენტრი): კორპუსი მოიცავს ა. ჭანტურიას მიერ თარგმნილი ლიტერატურის სრულ კატალოგს – როგორც კონფერენციის თეზისებს, ისე სამეცნიერო სტატიებსა და წიგნებს: საქართველოს მეცნიერებათა აკადემიის მოამბე, ინგლისურ-ქართული თეზისები; დმანისისა და ვანის არქეოლოგიური გათხრები (სტატიები და თეზისები); სამეროვანი გამოცემა პირველ ქართულ ნაბეჭდ წიგნებზე; სამეცნიერო ორენოვანი ჟურნალი „ქართველოლოგი"; ბ. ჯორბენაძის წიგნი „ქართველური ენები და დიალექტები" (1991), ე. ხინთიბიძის საერთო რედაქტორობით გამოცემული წიგნი „ქართული ლიტერატურა ევროპულ მეცნიერებაში" (2003); ე. შევარდნაძის წიგნი „დიდი აბრეშუმის გზა" (1999); გ. კალანდიას წიგნი „დადიანების სასახლეების საგანძური" (2003); ნ. ლომოურის „საქართველოს ისტორია" (1993) და მრავალი სხვა[114]. არიანე ჭანტურიას თარგმანების კატალოგი ასამდე დასახელების ერთეულისაგან შედგება და 5 მილიონამდე სიტყვაფორმას მოიცავს. ლექსიკონი სტრუქტურირებულია ტექსტური წყვილების პრინციპზე, რაც აადვილებს კორპუსის მონაცემთა ბაზაში სასურველი ქართული ან ინგლისური სიტყვის ან შესიტყვების ძიებას და ფასდაუდებელ პროგრამულ რესურსს წარმოადგენს ორენოვანი ლექსიკონისათვის ფუნქციური ეკვივალენტების დადგენის თვალსაზრისით.

–**ქართლის ცხოვრების პარალელური კორპუსი** (ილიას სახელმწიფო უნივერსიტეტის ლინგვისტურ კვლევათა ინსტიტუტი, პროექტის ავტორი პროფ. ნინო დობორჯგინიძე).

---

[113] იხ. LOEWE Text collection: NT Novum Testamentum georgice (http://titus.uni-frankfurt.de/texte/etca/cauc/ageo/nt/ntkpl/ntkpl.htm). პროექტის შესახებ დაწვრილებით იხ. ზემოთ, გვ. 75.

[114] იხ. თ. მარგალიტაძე, ევროპულ-ქართული პარალელური კორპუსები ქართული ლექსიკოგრაფიისა და მთარმნელობითი საქმიანობისათვის. https://margaliti.com/paralelkorpus.pdf



ქართლის ცხოვრების პარალელური კორპუსი მოიცავს: ა) ტექსტურ ბლოკს, რომელშიც წარმოდგენილია როგორც ნაბეჭდი, ასევე ხელნაწერი ტექსტები, რომელთა დიდი ნაწილი დღემდე არც ერთ გამოცემაში არ ყოფილა გათვალისწინებული. ამავე ბლოკში წარმოდგენილია ქართლის ცხოვრების ძველი სომხური თარგმანი და ამ უკანასკნელის ახალი ქართული თარგმანი. კორპუსში ძებნის განხორციელება შესაძლებელია სხვადასხვა ფილტრზე დაყრდნობით, და ბ) ვიზუალური მემკვიდრეობის ბლოკს, რომელშიც განთავსებულია ქართლის ცხოვრების შემცველ ხელნაწერებში დაცული მინიატურების, ხელნაწერთა ყდებისა და კალიგრაფიის ნიმუშები[115];

– **„ვეფხისტყაოსნის“ პარალელური კორპუსი** (ილიას სახელმწიფო უნივერსიტეტი, პროექტის ავტორი პროფ. ნინო დობორჯგინიძე) – კორპუსი მოიცავს ქართული ლიტერატურის ამ უმნიშვნელოვანესი ძეგლის სხვადასხვა ხელნაწერის სრულ და ფრაგმენტულ ნუსხებს; პარალელური კორპუსი ასევე შეიცავს „ვეფხისტყაოსნის“ ინგლისური თარგმანის სამ ვერსიას. ტექსტური რესურსის გარდა კორპუსი დამატებით მოიცავს მინიატურებსა და ილუსტრაციებს პოემის ბეჭდური და ხელნაწერი ვერსიებიდან.

[115] დაწვრილებით იხ. ნინო დობორჯგინიძე, ირინა ლობჟანიძე, ქართლის ცხოვრების პარალელური კორპუსი, მიმოხილვა.file:///C:/Users/Manana/AppData/Local/Temp/233-Article%20Text-467-1-10-20170101.pdf

– **„ვეფხისტყაოსნის“ თარგმანების პარალელური კორპუსი** (ფრანკფურტის უნივერსიტეტის ემპირიული ენათმეცნიერების ინსტიტუტი, პროექტის ავტორი პროფ. მანანა თანდაშვილი) – კორპუსი მიზნად ისახავს „ვეფხისტყაოსნის“ თარგმანების დიგიტალიზაციას 55 ენაზე თარგმანის სტრატეგიებისა და ეკვივალენტობის ავტომატური კვლევის მიზნით. დღეისათვის შექმნილია პარალელური კორპუსის ტექნოლოგიური ჩარჩო – კორპუსის პროტოტიპული ვერსია, რომელიც მოიცავს 13 ენაზე შესრულებულ თარგმანებს დაპარალელებულს ქართულ ორიგინალთან.



http://titus.uni-frankfurt.de/texte/caucasica/rustveli/rgd.html

პროექტის მიზნებიდან გამომდინარე, რომლის შესაბამისადაც შეიქმნა კორპუსის სტრუქტურული დიზაინი და სამომხმარებლო ინტერფეისი, „ვეფხისტყაოსნის“ თარგმანების პარალელური კორპუსი ქართულ სამეცნიერო სივრცეში მრავალენოვანი პარალელური კორპუსის შექმნის პირველი მასშტაბური მცდელობაა.

– **პოლიტიკური ტექსტების ქართული მთარგმნელობითი კორპუსი** (ბათუმის შოთა რუსთაველის სახელმწიფო უნივერსიტეტი, პროექტის ავტორი პროფ. ხათუნა ბერიძე). პროექტი მოიცავს პოლიტიკური მეცნიერების, ანალიტიკურ, საგაზეთო, საკანონმდებლო და სხვა ჟანრის ტექსტებს და მოიცავს: პოლიტიკური ტექსტებისა და კონცეპტების მონოლინგვურ განმარტებით ლესიკონს, კონტექსტური მონაცემების ბაზებს, რომლებიც შედგება კომპარატიული და პარალელური ტექსტების სექციებისა და ქვესექციებისაგან[116].

---

[116] ხათუნა ბერიძე, გრიგოლ კახიანი, პოლიტიკური ტექსტების ქართული მთარგმნელობითი კორპუსი, საერთაშორისო კონფერენცია „ქართული ენა და თანამედროვე ტექნოლოგიები“, კონფერენციის თეზისები, თბილისი 2013.

ე) სპეციალური ანუ თემატური კორპუსი

სპეციალურ ან თემატურ კორპუსებად მიჩნეულია ისეთი კორპუსები, რომელიც ენობრივ რესურსებს აერთიანებს გარკვეული სპეციალური თემატიკის მიხედვით. თემატური კორპუსია, მაგალითად, დეფექტური მეტყველების კორპუსი (Fromkin's Speech Error Database), ბავშვთა მეტყველების კორპუსი (Child Language Data Exchange System), ბონის პრესის კორპუსი (Bonner Zeitungs-Korpus), დორტმუნდის ნატის კორპუსი და სხვ.

მაგალითისათვის განვიხილოთ თემატური კორპუსის ერთი შემთხვევა, კერძოდ ალპინისტური კორპუსი – **Corpus: Text+Berg digital.**

თემატური კორპუსი Text+Berg digital (ე.წ. ალპინისტური კორპუსი) შეიქმნა დრეზდენის ტექნიკურ უნივერსიტეტში დოქტ. ნოა ბუბენჰოფერის (Noah Bubenhofer) და ციურიხის უნივერსიტეტის პროფ. მარტინ ფოლკის (Martin Volk) მიერ სამეცნიერო პროექტის ფარგლებში და მიზნად ისახავდა ალპინისტური ლიტერატურის დიგიტალიზაციას თემატური კორპუსის შექმნის მიზნით.

კორპუსში შევიდა შვეიცარიის ალპური კლუბების წლიური ანგარიშების ბეჭდური რესურსები – 1864-1923 წლებში ყოველწლიური გამოცემები, 1923 წლიდან კი ყოველთვიური ჟურნალი „ალპები". კორპუსის მოცულობა შეადგენს 37 მილიონამდე ტოკენს.

ალპინისტური კორპუსის, როგორც თემატური კორპუსის შექმნა, გარდა ისტორიული ფაქტების კვლევისა, ლინგვისტური თვალსაზრისითაც იყო მნიშვნელოვანი. აქ მოვიტანთ რამდენიმე საინტერესო დაკვირვებას კორპუსპრაგმატული თვალსაზრისით:

- **თხრობის შინაარსობრივი და სტრუქტურული მაჩვენებელი** – ადრეულ ტექსტებში ჭარბობს ნარატიული თხრობა, წინადადებები გრძელია. გვიანდელ ტექსტებში წინადადებები მოკლდება, სამაგიეროდ მატულობს პასიური კონსტრუქციები და ნარატიული თხრობა იცვლება ექსპრესიული თხრობით, რომელიც ძირითადად გადმოცემულია პირველ პირში. თხრობის შინაარსობრივი და სტრუქტურული მაჩვენებლები იზომება, მაგალითად, პასიური კონსტრუქციების, ექსპრესიულობის გამომხატველი ლექსიკური და ფუნქციური ელემენტების პოვნიერების გზით.
- ტექსტის შეფასების ერთ-ერთ პარამეტრს წარმოადგენს **ტექსტის ლექსიკური სიმდიდრის პარამეტრი**, რომელიც იზომება ტექსტში ლექსემების რაოდენობის მიხედვით: რაც უფრო ხშირად გეხვდება ერთი და იგივე სიტყვა ტექსტში, მით უფრო დაბალია აღნიშნული პარამეტრი და შესაბამისად, ტექსტი – უფ-

რო პრიმიტი'ული და მარტივად გასაგები. როგორც კორპუსის ავტორის, ნოა ბუბენჰოფერის კვლევის შედეგებიდან ჩანს, ლექსიკური სიმდიდრის პარამეტრთან ერთად წინადადების სიგრძე და ტექსტის სირთულე საგრძნობლად ეცემა 1864-1960 წლებში, 1960 წლიდან კი ისევ მატულობს.

- თანამედროვე ტექსტებში შედარებით ჭარბობენ ლექსემები ტელეფონი, ფოტო, მარშრუტი, პარასკევი, ინფორმაცია, სტატია, თემა, სპორტის სახეობები, რომელიც გვიანდელ ტექსტებში დასტურდება, გაცილებით მრავალფეროვანია, ვიდრე ადრეულ ტექსტებში.
- გვიანდელ ტექსტების ლექსიკურ მარაგში აღარ გვხვდება განცდისმიერი ლექსიკა, რომანტიული, სენტიმენტალური, ექსპრესიული შინაარსის გადმომცემი სიტყვები, რაც ამ პერიოდის ალპინისტებში სამყაროს არარომანტიკული აღქმის დამადასტურებელია. ადრინდელ ტექსტებში „ემოციურობის" კოეფიციენტის გაზომვა, გარდა ლექსიკური საზომისა (შესაბამისი ლექსემების პოვნიერება, როგორიცაა მშვენიერი, განსაცვიფრებელი, შთამბეჭდავი) კომპლექსური n-gramm – ების მეშვეობითაც დასტურდება. მაგ.: ელიპტური ძახილის წინადადებები ADJ+NN+! (Ein wunderbares Wandergebiet! ქართულში ამ ტიპის ელიფსურ წინადადებებს შეესაბამება არაელიფსური წინადადებები კოპულით: „(რა) მშვენიერი დღეა!"). გვიანდელი ტექსტების ენა რეპორტაჟული ტიპისაა და აღარ არის დაშტამპული პიროვნული შთაბეჭდილებებით, მოზომილი და რაციონალურია.
- ადრეულ ტექსტებში პირის ნაცვალსახელების გვერდით დომინირებს ანთროპონიმები ან სპეციფიკური ლექსიკა (მეგობარი, თანაგუნდელი). გვიანდელ ტექსტებში იგივე სურათი გვაქვს, რაც ალპინისტების გუნდურობისა და სოციალიზაციის ძირითად მახასიათებელ ნიშნად რჩება.

როგორც ალპინისტური კორპუსის კვლევები გვიჩვენებს, საუკუნეხახევრის განმავლობაში საგრძნობლად შეიცვალა მთამსვლელთა მოტივაცია: თუ მე-18 საუკუნის მეორე ნახევარში მთამსვლელთა ალპებში მოგზაურობის ძირითადი მიზანი სამეცნიერო-შემეცნებითი იყო, მე-20 საუკუნის 30-იან წლებში წინა პლანზე ჩანს არა ბუნება მთელი თავისი მრავალფეროვნებით, არამედ თავად ადამიანი – ერთი მხრივ პიროვნული EGO-ს და მეორე მხრივ, ტექნოკრატიული მსოფლიოს სურათი. გაცილებით ძლიერდება თანამედროვე ალპინისტურ ლიტერატურაში.

კორპუსში აღნიშნული კვლევის განხორციელება შესაძლებელი გახდა იმდენად, რამდენადაც კორპუსი სრულად არის ანოტირებული, რამაც, თავის მხრივ, განსაზღვრა კვლევების ეფექტიანობა.

ფრანკფურტის უნივერსიტეტში დიგიტალური ქართველოლოგიის ფარგლებში შექმნილი თემატური კორპუსის საუკეთესო მაგალითია **ქართული პოლიტიკური ტექსტების კორპუსი** (მოიცავს საქართველოს პრეზიდენტების გამოსვლების დისკურსს, აგრეთვე ქართველ პოლიტიკოსთა გამოსვლებს პარლამენტის სხდომებსა და თოქ-შოუებში) და **იურიდიული ტექსტების კორპუსი** (მოიცავს საშუალი ქართულის იურიდიულ ტექსტებს).

5.3. კორპუსის ტიპები

კორპუსში განთავსებული მასალების მიხედვით განასხვავებენ სხვადასხვა ტიპის კორპუსებს. ქვემოთ განვიხილავთ სამი ტიპის კორპუსს – ისტორიულ კორპუსს, ანოტირებულ კორპუსს და ეროვნულ კორპუსს.

ა) ისტორიული კორპუსი

ისტორიული კორპუსი (historical corpus) წარმოადგენს კორპუსის ტიპს, რომელიც მოიცავს მოცემული კონკრეტული ენის ან ენათა ისტორიული არსებობის პერიოდში შექმნილ დოკუმენტთა ფართო ჟანრობრივ რეპერტუარს და დიაქრონიული კვლევის საშუალებას იძლევა. კვლევის ამ სპეციფიკური შესაძლებლობიდან გამომდინარე, ისტორიულ კორპუსს ხშირად დიაქრონულ კორპუსსაც უწოდებენ.

ისტორიული კორპუსი შეიძლება რომელიმე ქრონოლოგიურ პერიოდზე იყოს შეზღუდული. მაგ. ინგლისური ენის ისტორიული კორპუსი ARCHER (A Representative Corpus of Historical English Registers) მოიცავს 1650-1999 წლების ინგლისური ენის მასალას (მოცულობა: 1,8 მილიონი ტოკენი), მაშინ როდესაც ინგლისური ენისათვის შექმნილი მეორე კორპუსი Penn Parsed Corpora of Historical English გაცილებით ადრეულ მასალებს შეიცავს (1150-დან 1914-ის ჩათვლით) და მოცულობითაც თითქმის ორჯერ აღემატება პირველს – მისი მოცულობა: 3.9 მილიონ ტოკენს აღწევს.

ქართული ენის ეროვნული კორპუსი სწორედ ისტორიული კორპუსების ეტაპობრივი განვითარებით შეიქმნა და მოიცავს შემდეგ ქვეკორპუსებს: ძველი ქართული ენის კორპუსს, საშუალი ქართული ენის კორპუსს და ახალი ქართული ენის კორპუსს.

ბ) ანოტირებული კორპუსი

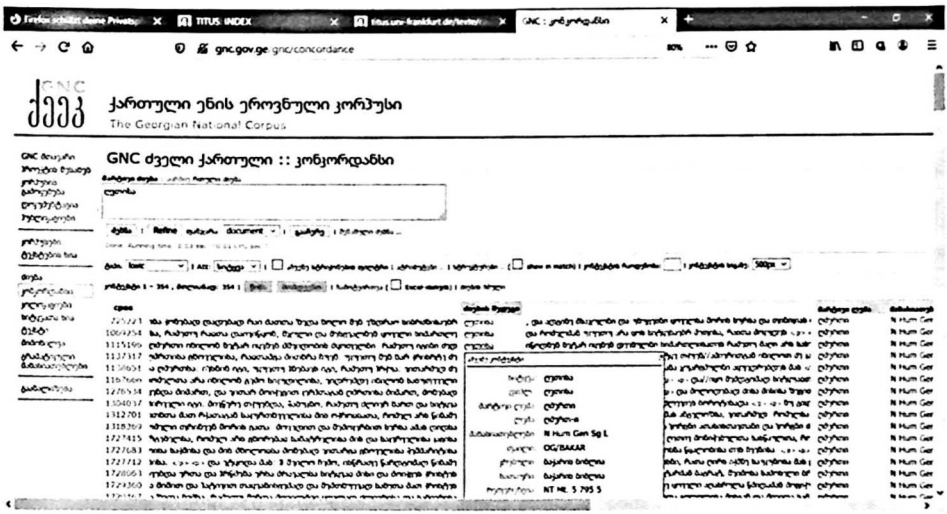
ანოტირებულ კორპუსად მიიჩნევა ისეთი კორპუსი, რომელიც კორპუსის საგანგებო კომპონენტით – ანოტაციით არის აღჭურვილი.

კორპუსლინგვისტიკის განვითარების ერთ-ერთ უმნიშვნელოვანეს ეტაპად მიჩნეულია ლინგვისტური ანოტაციის პროცესის პროგრამული ავტომატიზაცია. რო-

გორც ზემოთ უკვე აღვნიშნეთ, ამ მიზნის მისაღწევად კორპუსის პროგრამულ ინფრასტრუქტურაში ჩაშენებულია სპეციალური ინსტრუმენტები, რომელთა საშუალებითაც ხორციელდება მიმართება სხვადასხვა სახის პირველად მონაცემთა ბაზებთან, რაც ლექსემათა ავტომატური ანოტირების საშუალებას იძლევა. ამგვარი მექანიზმით აღჭურვილ კორპუსს ანოტირებული კორპუსი ეწოდება.

ანოტირებული კორპუსის შექმნის დროს ერთ-ერთ მნიშვნელოვან საკითხს წარმოადგენს ტექსტის ასახვის პრობლემა. უპირველეს ყოვლისა, უნდა შემუშავდეს ტექსტის კორპუსში ასახვის სტრატეგია – **დიპლომატიური ტექსტი** შედის კორპუსში, როგორც კორპუსის რესურსი თუ **ნორმატიული ტექსტი**. ტექსტის კორპუსში ასახვის საუკეთესო ვარიანტია ტექსტის ინტეგრაცია კორპუსში მრავალდონიანი ანოტაციის სახით. ასეთ შემთხვევაში ტექსტი რამდენიმე დონეზეა განფენილია და შეიცავს როგორც დიპლომატიურ, ისე ნორმატიულ ტექსტს.

ქართული ენის ეროვნული კორპუსი ანოტირებულ კორპუსს წარმოადგენს. გარდა ტექსტისა, იგი შეიცავს როგორც გარემეტამონაცემებს (ავტორი, წყარო, თარიღი), ისე გრამატიკულ მახასიათებლებს:



ქართული ენის ეროვნული კორპუსის ანოტაციის სახეებს დაწვრილებით განვიხილავთ ქვემოთ (იხ. ქვეთავი **ქართული ენის ეროვნული კორპუსი**)

გ) ეროვნული 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**. დღეისათვის ენის ეროვნული კროპუსის მნიშვნელობა სცილდება მეცნიერების ფარგლებს, სახელმწიფო მნიშვნელობის რანგში არის აყვანილი და განიხილება სახელმწიფო ატრიბუტიკის – გერბის, დროშისა და ჰიმნის გვერდით.

ენის კორპუსსა და **ენის ეროვნულ კორპუსს** ერთადერთი საერთო ნიშანი გააჩნია: 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.

ქართული ენის მაგალითზე თუ 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** და ენის კვლევა თანამედროვე ტექნოლოგიების გამოყენებით;

3. მისი აქტუალობა მდგომარეობს **კვლევის ეფექტიანობაში** (საკვლევი ველის მოცულობა და ძიების სისწრაფე) და გამოირჩევა **ვერიფიკაციის მაღალი ხარისხით** (სულ რამდენიმეწუთიანი ოპერაციის შედეგად შესაძლებელია ნებისმიერი თეორიული პოსტულატის დადასტურება (ვერიფიცირება) ან უარყოფა (ფალსიფიცირება).

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:

- ეფექტიანი ენობრივი პოლიტიკის გატარების ხელშეწყობა;
- ქვეყნის საგანმანათლებლო სისტემის მოდერნიზაციის ხელშეწყობა;
- 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 იმაში, რომ იგი უზრუნველყოფს ქართული ენის, როგორც სახელმწიფო ენის, სრულფასოვან და წარმატებულ 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. ახალი ზელანდია, რომლის სახელმწიფო ენას ადგილობრივი, პოლინეზიური მაორი-ენის გვერდით ინგლისური ენა წარმოადგენს, ვერცერთ აქ აღნიშნულ კორპუსს ვერ გამოიყენებს სახელმწიფო ინტერესებისათვის, ვინაიდან ნებისმიერი ეროვნული კორპუსი მხოლოდ მოცემული ქვეყნისათვის შეიძლება იყოს სახელმწიფო ენის სრულფასოვანი ფუნქციონირების გარანტი და მოქნილი იარაღი.

ეროვნული კორპუსების აგებულება ძირითადად მსგავსია, ისინი ერთმანეთისაგან მოცულობის თვალსაზრისით განსხვავდებიან.

ქართული ენის ეროვნული კორპუსი დღეისათვის 202 მილიონი ტოკენის მოცულობისაა. მომდევნო თავში დაწვრილებით განვიხილავთ ქართული ენის ეროვნული კორპუსის შექმნის ისტორიას, კორპუსის სტრუქტურასა და მის არსებით მახასიათებლებს.

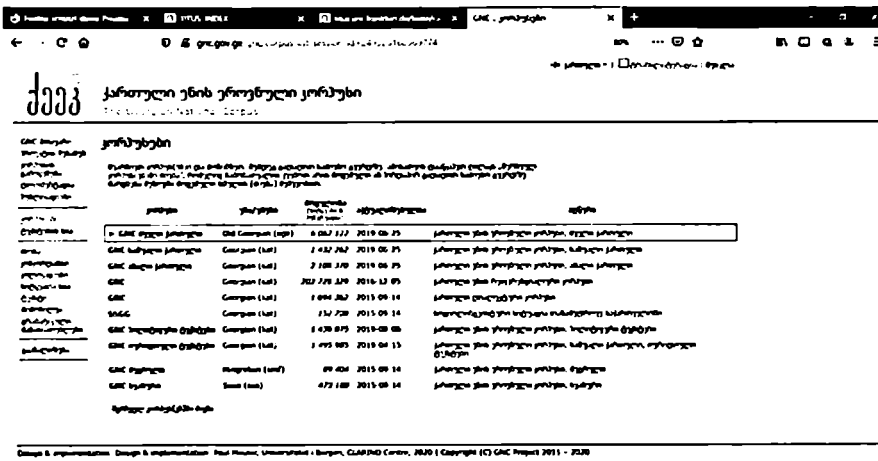
5.4. კორპუსის დაბალანსება

დაბალანსება (**balancing**) კორპუსის ერთ-ერთი თვისებრივი მახასიათებელია და გულისხმობს კორპუსში ჩართული რესურსების ჟანრული მრავალფეროვნების გაწონასწორებას. დაბალანსება შესაძლებელია როგორც სინქრონიული, ისე დიაქრონიული თვალსაზრისით. დიაქრონიული კვლევებისათვის მხოლოდ ისეთი კორპუსების გამოყენებაა მიზანშეწონილი, რომელიც დიაქრონიულად დაბალანსებულია. ასეთ კორპუსებს მიეკუთვნებიან, მაგ., ისტორიული კორპუსები.

სინქრონიულად დაბალანსებული კორპუსები ძირითადად უნდა მოიცავდეს ენის სხვადასხვა რეგისტრებს, უპირველეს ყოვლისა, დიალექტებსა და სოციოლექტებს. ენის ემპირიული ფორმის ყველა შესაძლო ვარიანტის ასახვა კორპუსში უზრუნველყოფს დაბალანსების ხარისხს. ემპირიულ ფორმებში იგულისხმება არა მარტო წერილობითი, არამედ ზეპირმეტყველების რეგისტრიც. ამ რეგისტრის შიგნით მრავალფეროვნება მიიღწევა, მაგ., ასაკობრივი ჯგუფების მრავალფეროვნებით (ბავშვთა მეტყველება, მოზარდთა მეტყველება, ასაკოვანთა მეტყველება), პროფესიული

ჯგუფების მრავალფეროვნებით (მასწავლებელთა მეტყველება, ექიმთა მეტყველება, მძღოლთა მეტყველება), სოციალური ჯგუფების მრავალფეროვნების მიხედვით (ე.წ. აკადემიური მეტყველება, ყოველდღიური მეტყველება, ქუჩის მეტყველება), სიტუაციების მრავალფეროვნების მიხედვით (თამადის მეტყველება, კონფლიქტური სიტუაციების დროს მეტყველება). დაბალანსებულ კორპუსში ცალკე აისახება, მაგალითად, ბავშვის დეფექტური მეტყველება, ალკოჰოლის ზემოქმედების ქვეშ მყოფი ადამიანის მეტყველება, სპეციფიური დიალოგები (მაგ. კომერციული საბანკო ან ყოველდღიური სავაჭრო დიალოგი) და ასე შემდეგ.

დღემდე არსებული კორპუსები, მათი ტიპებისა და სახეებიდან გამომდინარე, არსებითად განსხვავდებიან ერთმანეთისაგან დაბალანსების თვალსაზრისით. მაგ., თემატური კორპუსები ყველაზე ნაკლებად არიან ბალანსირებული, ხოლო ეროვნული კორპუსები, როგორც წესი, დაბალანსების ყველაზე მაღალი მაჩვენებლებით ხასიათდებიან.



ქართული ენის ეროვნული კორპუსი ამჟამად დაბალანსების ეტაპზე იმყოფება. კორპუსს ბოლო ხანს ორი თემატური ქვეკორპუსი შეემატა: პოლიტიკური ტექსტების კორპუსი და იურიდიული ტექსტების კორპუსი.

კორპუსის დაბალანსება თავისთავად გულისხმობს კორპუსის განერცობას სხვადასხვა პარამეტრების მიხედვით, მაგალითად, აღნიშნული თემატური კორპუსებიდან პირველი – პოლიტიკური ტექსტების კორპუსი უნდა განვიხილოთ, როგორც თემატური კორპუსი, რომელიც მხოლოდ ახალი ქართულის ენობრივ სისტემას ასახავს, მაშინ როდესაც იურიდიული ტექსტების კორპუსი საშუალი ქართულის თემატურ კორპუსს წარმოადგენს.

6. კორპუსის სტრუქტურა

კორპუსის სტრუქტურას შეადგენს სამი არსებითი კომპონენტი: **პირველადი მონაცემი, მეტამონაცემი და კორპუსმენეჯერი.** განვიხილოთ თითოეული ცალ-ცალკე.

6.1. სტრუქტურირებული ტექსტი კორპუსში

კორპუსის უმცირეს შინაარსობრივ საბაზისო კომპონენტს წარმოადგენს პირველადი მონაცემი – კოჰერენტული ტექსტი. კორპუსის ეს უმთავრესი კომპონენტი აუცილებლად უნდა იყოს სტრუქტურულად დამუშავებული და ზუსტად უნდა ასახავდეს დედანს. განვიხილოთ ეს საკითხი ქართული ლიტერატურის უძველესი ძეგლის, „წამებაი წმიდისა შუშანიკისი“-ს მაგალითზე.

ტექსტი, რომელიც ქვემოთ არის წარმოდგენილი, გახლავთ „შუშანიკის წამების“ ელექტრონული ვერსია:

II. და ოქტმა მეფისაგან სპარსთაისა პიტიახშმან. და ვითარცა მოიწია ოგი საზღვართა ქართლისათა, ქცეულმან მას ჴერეთისასა, ზარახვა-ყო, რაითა აღწყოს და ნიშამიდგებმენ მას აზნაურნი და მისნი ძენი და მსახურნი მისნი, რაითა მათ გამო, ვითარცა გველი, სოფლად შეიდეს. და წარმოსძართა საიდასაბოთა ცხენოსა მოსაჲ თჳსი. და ვითარცა მოვიდა დასპა, რომელსა სახელი პრქუან ცურტავ.

და ვითარცა მოიწია და შეიდა წიხაშე შუშანიკისა, დედოფლისა ჩუენისა, კითხვაი პრქუა, ხოლო სპასტრელმან შუშანიკ ვითარცა ნიშაისწარმეტყველმან პრქუა: „ცუღეთე სულითა ცხოველ არს ოგი, ცოცხალმცა ხართ, ოგიც და შენცა. ცუღეთე სულითა მოჰკუდართ ხართ, მოკითხვაი ჳგე შენი შენდაჳე მოიქციო!“. ხოლო კაცმან მას ჳერ იკადრა სიტყჳად მისა. ხოლო წმიდაი შუშანიკ აფუცებდა და ჰკითხჳდა დამტკიცებულად. ხოლო კაცმან უთხრა მართალი და პრქუა, ვითარშედ: „ვარსქენ უგარ-ყო ჰეშმარიტი ღმერთი“.

ვითარცა ესმა წეტარსა შუშანიკს, დავარდა ქუეყანასა ზედა და თავსა დამართ სცემდა და ცრემლითა მწარითა იტყოდა: „საწყალებელ იქმნა უბადრეკი ვარსქენ, რამეთე უგარ-ყო ჰეშმარიტი ღმერთი და აღარა ატროშანი და შჳერთო ოგი უღმრთოთა“.

და აღგა და დაუტევა ტაძარი თჳსი და ღმრთისმოშიშებით ეკლესიად შჳდა და თაჱს მიუვლია სამწი ოგი ძენი და ერთი ასული და წარადგინნა ოგიი ნიშაზე საჯერთხეველისა და ესრეთ ილოცვიდა და იტყოდა: „ეჳხილ ღმერთი, შენსნი მოცემულნი არიან და შენ დაიცჳენ ესენი, განათლებულნი წმიდისაგან ემბაზისა, მადლითა სულისა წმიდისათა, ყოფად ერთ სამწყსო ერთისა მწყემსისა უფლისა ჩუენისა იესე ქრისტესა!“

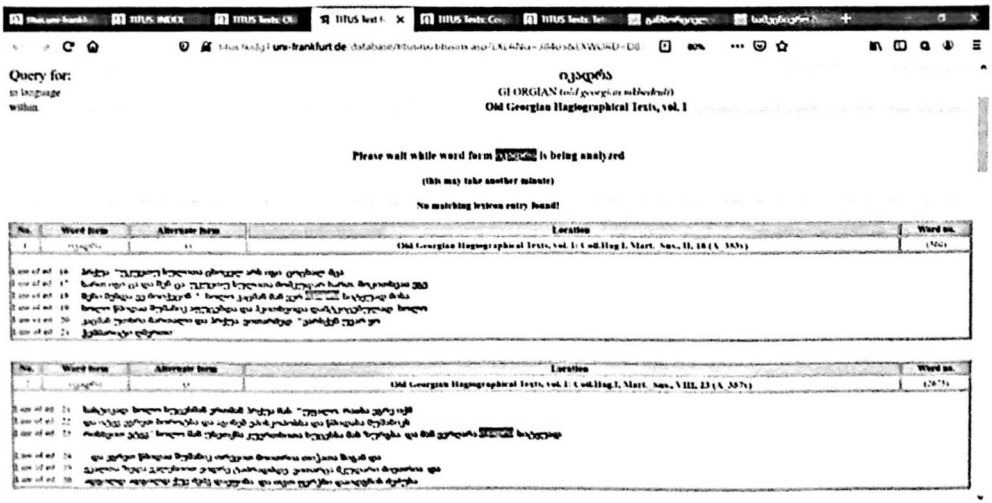
და ვითარცა მწუხრისა ჟამი აღსრულეს, სახლაკი ერთი მცირე პოვა მახლობელად ეკლესიასა და შეიდა მუნ მიწა სიმწარითა სავსე და მიეყარდნა ყურესა ერთსა და მწარითა ცრემლითა ტირდა.

III. ხოლო ეპისკოპოსი ოგი სახლისა მის პიტიახშისაჲ, რომელსა სახელი ერქუა აფრე, არა მუნ იყო, არამედ კაცისა ვისამე წმიდისა ეპისკოპოსისა მას. და მყის მოიწია ჩუენდა დიაკონი მიზეთი და გკითხა ჩუენ ესე ყოველი – მოსლუაჲ პიტიახშისაჲ და საქმენი დედოფლისანი. ხოლო ჩუენ აღვივსებით მწუხარებითა და დიდითაი ტირილითა ვტირდით განშორებულნი ცოდვათ ჩუენთა წარმოჩინებებისათჳს.

როგორც ეხედავთ, ელექტრონული ვერსია წარმოადგენს „სუფთა“ ტექსტს ყოველგვარი დამუშავების გარეშე და ხელნაწერი ან ნაბეჭდი ვერსიისაგან მხოლოდ იმით განსხვავდება, რომ ეს უკანასკნელი ელექტრონულ მატარებლებზე არის გადატანილი. იგივე ტექსტის სტრუქტურირებული ვერსია ამგვარად გამოიყურება TITUS-ში:



ამ ორ ტექსტს შორის სხვაობა იმაში მდგომარეობს, რომ ეს უკანასკნელი ორიგინალის ტექსტის გადამუშავებულ, სტრუქტურირებულ ვერსიას წარმოადგენს: მასში ზუსტად არის მითითებული ძეგლის შიდა სტრუქტურა (სათაური, თავი პირველი, თავი მეორე და ა.შ.), ელექტრონულ ტექსტში სტრიქონებისა და გვერდების განლაგება ზუსტად ემთხვევა საყრდენ ხელნაწერში ტექსტის განფენას. შესაბამისად, შუშანიკის წამების TITUS-ისეული ვერსია ორიგინალის ვირტუალურ ადეკვატს წარმოადგენს.



გარდა ამისა, ტექსტში დამატებით არის მარკირებული არსებითი სახელის სპეციფიკური ჯგუფები – ტოპონიმები, ანთროპონიმები და ჰიდრონიმები, რისთვისაც ელექტრონულ ტექსტში გამოყენებულია ტექსტის მარკირების საგანგებო ნიშნულები (Extensible Markup).

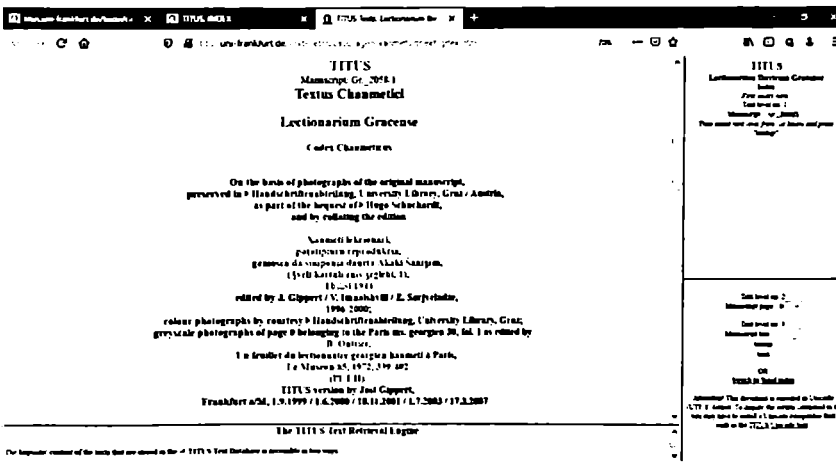
ტექსტი დამუშავებულია ტექსტური მარკირების ყველაზე გავრცელებული ენის XML (Extensible Markup Language) ფორმატში. ტექსტზე დართული საძიებო სისტემის მეშვეობით რომელიმე სიტყვაზე (მაგ.: *იკადრა*) დაწკაპუნებით კი იძებნება აღნიშნული სიტყვის გამოყენების ყველა კონტექსტი მოცემულ ტექსტში.

6.2. მეტამონაცემი

კორპუსის მეორე უმნიშვნელოვანეს კომპონენტს წარმოადგენს მეტამონაცემი – ანუ მონაცემი ენობრივი მონაცემის შესახებ. მეტამონაცემი შეიცავს მეორად, განმარტებულ ინფორმაციას პირველადი ენობრივი მონაცემის შესახებ, როგორიცაა მაგ.:

- პირველადი მონაცემის (მაგ., წარწერის, ხელნაწერის, ნაბეჭდი წიგნის) შექმნის ადგილი;
- პირველადი მონაცემის ტექსტის ავტორის ვინაობა;
- პირველად მონაცემთან დაკავშირებული რელევანტური თარიღები (ტექსტის დაწერის, გადაწერის).
- პირველადი მონაცემის (ანუ ტექსტის) დიგიტალური ვერსიის შემქმნელი, ადგილი და დრო და ასე შემდეგ.

მეტამონაცემი შესაძლოა თან ერთვოდეს პირველად მონაცემს – ტექსტს (იხ. TITUS) – ან განთავსდეს სპეციალური ფორმატის (header) დამოუკიდებელ ბაზაში.



კორპუსების შექმნის ადრეულ პერიოდში, გასული საუკუნის 80-იან წლებში, როდესაც ჯერ კიდევ არ იყო დამუშავებული მეტაინფორმაციებისათვის საგანგებო სტრუქტურული დანართის პრინციპი და კრიტერიუმები, მეტამონაცემი თავად ჰქონდა დართული ტექსტს ერთგვარი ქუდის სახით, როგორც მაგ. ზემომოყვანილ TITUS-ის შემთხვევაში გვაქვს. როგორც აქ მოტანილი მეტამონაცემებიდან ჩანს, TITUS-ის ეს ტექსტი ეყრდნობა ნაშრომს, რომელიც 1944 წელს გამოსცა აკაკი შანიძემ ძველი ქართული ენის ძეგლების პირველ ტომში და რომელიც ხანმეტი ლექციონარის ფოტოტიპურ რეპროდუქციას წარმოადგენს.

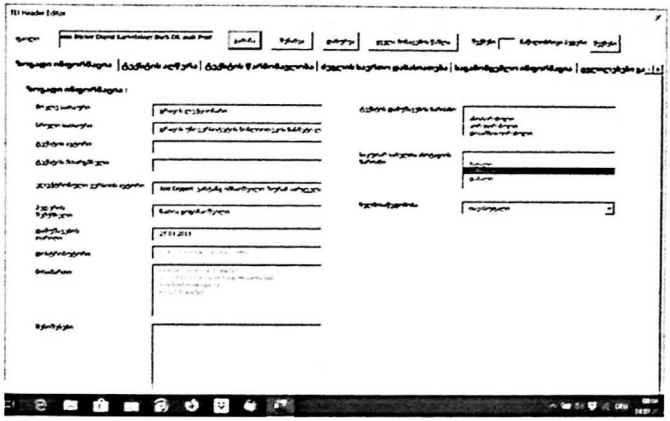
ჰედერში მოცემულია აგრეთვე ინფორმაცია ელექტრონული ვერსიის შემქმნელთა (ამ კონკრეტულ შემთხვევაში ი. გიპერტის, ე. იმნაიშვილის და ზ. სარჯველაძის) შესახებ. აქვეა მოცემული ინფორმაცია იმის შესახებ, თუ სად ინახება დედანი (ხელნაწერის ძირითადი ნაწილი ინახება გრაცის უნივერსიტეტის ბიბლიოთეკის ხელნაწერთა განყოფილებაში, ხელნაწერის პირველი გვერდი კი პარიზის ქართული ხელნაწერების ფონდშია დაცული და აღწერილია და გამოცემული ბერნარდ უტიეს მიერ 1972 წელს).

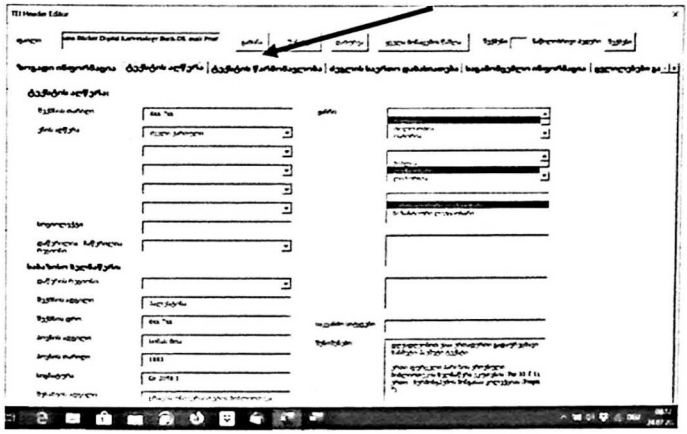
დღესდღეობით შემუშავებულია Header-ის ახალი ვერსია, რომელიც Header-ის აგების საერთაშორისო კონსორციუმის TEI (The Text Encoding Initiative, შემოკლებით TEI) მიერ იქნა შემუშავებული და საერთაშორისო სტანდარტს წარმოადგენს[117]. აღნიშნული სტანდარტი ემსახურება დიგიტალური ტექსტების საერთაშორისო მიმოქცევაში ჩართვას და აადვილებს რესურსების გამოყენებას საერთაშორისო სამეცნიერო სივრცის ფარგლებში. მეტაინფორმაცია ჰედერში ბლოკებად არის ჩადებული: ზოგადი ინფორმაცია, ტექსტის აღწერა, ტექსტის წარმომაელობა, ძეგლის საერთო დახასიათება, საგამომცემლო ინფორმაცია, ცვლილებები (მეტამონაცემების ასასახად დაზუსტების ან განერცობის შემთხვევაში).

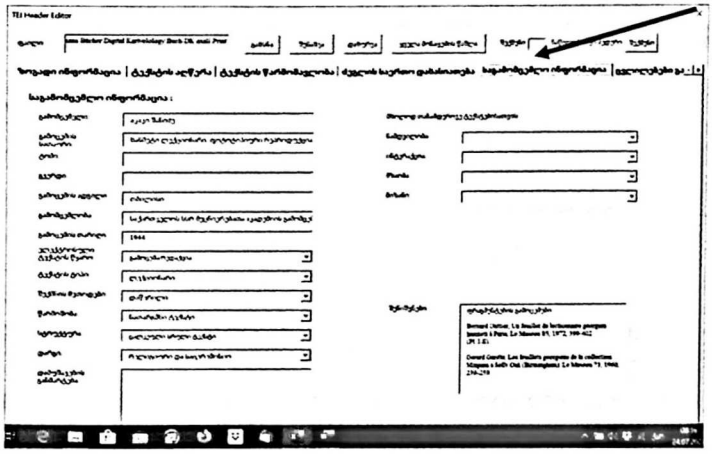
ზემოთ მოტანილი გრაცის ხანმეტი ლექციონარის ტექსტის ჰედერი ქართული ენის ეროვნულ კორპუსში ასე გამოიყურება:

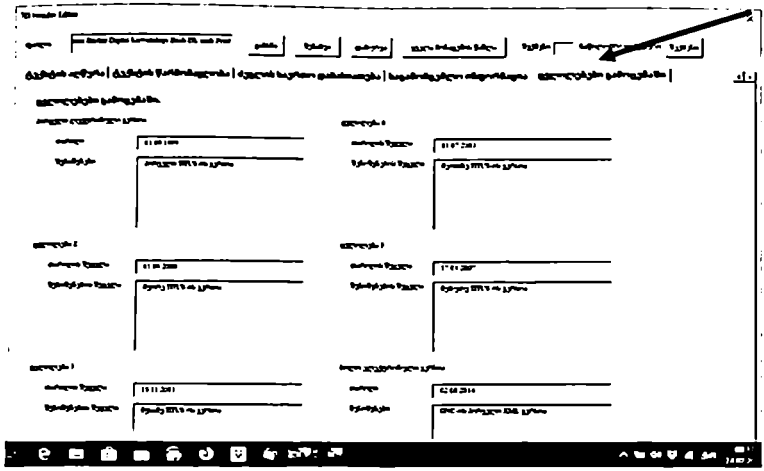
---

[117] The Text Encoding Initiative (TEI) is a consortium which collectively develops and maintains a standard for the representation of texts in digital form. Its chief deliverable is a set of Guidelines which specify encoding methods for machine-readable texts, chiefly in the humanities, social sciences and linguistics. Since 1994, the TEI Guidelines have been widely used by libraries, museums, publishers, and individual scholars to present texts for online research, teaching, and preservation. In addition to the Guidelines themselves, the Consortium provides a variety of resources and training events for learning TEI, information on projects using the TEI, a bibliography of TEI-related publications, and software developed for or adapted to the TEI.









6.3. კორპუსმენეჯერი

კორპუსის ერთ-ერთ ძირითად შემადგენელ კომპონენტს წარმოადგენს **კორპუსის მართვის მენეჯერი**. კორპუსმენეჯერი უსრუნველყოფს ძიების პროცესს, რომელიც საშუალებას იძლევა ძიება განვახორციელოთ ერთი ან რამდენიმე პარამეტრის მიხედვით.

მაგალითად:

ა) მორფოლოგიური ნიშნის მიხედვით:

მომიძებნე ყველა, მოთხრობით ბრუნვაში მდგარი არსებითი სახელი;

ბ) ჰედერში მოცემული მეტამონაცემების მიხედვით:

მომიძებნე ყველა რესურსი, რომლის ავტორიც არის ილია ჭავჭავაძე;

გ) ან ერთდროულად, ორივე სახის პარამეტრის მიხედვით:

მომიძებნე ყველა, მოთხრობით ბრუნვაში მდგარი სახელი იმ ტექსტებში, რომელიც ილია ჭავჭავაძის მიერაა შექმნილი.

კორპუს-საძიებო სისტემა, თავის მხრივ, შეიცავს კორპუსულ ძიებაზე მომართულ ექსპლორაციულ ალგორითმებს – საგანგებო ხელსაწყოს (ე.წ. tool-ს), რომელიც უსრუნველყოფს კორპუსში ძიების პროცესს.

ძიების დროს ტექსტის დანაწევრება ხდება ფრაგმენტებად, რომელიც მეორე ეტაპზე ელემენტებად (***n*-gram**) ჯგუფდება. ამგვარ დაჯგუფებას შესაძლოა საფუძვლად ედოს ასოების, მორფემების, ლექსემების მსგავსებები. ***n*-gram**-ის სახეებია: მონოგრამები, ბიგრამები, ტრიგრამები და ა.შ. გვხვდება მულტიგრამებიც შესამაბისად მაკონსტრუირებელი ნიშნის რაოდენობის მიხედვით.

***n*-gram**-ი ფორმალურად ასე განისაზღვრება: თუ **Σ** სასრული ანბანია და **n** წარმოადგენს მთლიან დადებით რიცხვს, მაშინ ***n*-gram**-ი არის **Σ** ანბანისაგან შემდგარი ერთი **w** სიტყვა **n** სიგრძით, ანუ **w=(w1,...,wn) ∈ Σn.**

იხ. მაგალითი ქართულისათვის:

Monogramm/Unigramm	1	ა
Bigramm	2	აი
Trigramm	3	სიო
Tetragramm	4	თათი
Pentagramm	5	ვაშლი
Hexagramm	6	თამასა
Heptagramm	7	თბილისი
Oktogramm	8	წიგნების
...	...	
Multigramm	N	კორპუსლინგვისტიკა

კორპუსული ძიება წარმოადგენს ***n*-gram**-ზე დაფუძნებულ მიზანმიმართულ ძიებას კორპუსში, რომლის ეფექტიანობა მიემართება კორპუსის ინჟინერიის პრერეკვიზიტებს ანუ წინასწარ დადგენილ პირობებს.

7. კორპუსის პრერეკვიზიტები

კორპუსი, როგორც წესი, აღჭურვილია საგანგებო პრერეკვიზიტებით, რომელიც კორპუსში ძიების პროცესს უზრუნველყოფს. კორპუსის პრერეკვიზიტებიდან გამომდინარე კორპუსში შესაძლებელია როგორც მარტივი, ისე რთული ძიების განხორციელება. მარტივი ძიება უნივერსალურ პრინციპს – გრაფემათა ლინეარულ თანმიმდევრობას ეყრდნობა, ან ძიებას განახორციელებს მორფოლოგიური პარამეტრების მიხედვით. შესაბამისად, მარტივი ძიების პრინციპი ყველა ენისათვის მისაღებია. რთული ძიება კი, ენობრივი სისტემის თავისებურებებიდან გამომდინარე, მეტნაკლებად განსხვავებულია და ყოველი ენის კორპუსისათვის ადაპტირებას მოითხოვს. განვიხილოთ რამდენიმე მაგალითი.

ა) მარტივი ძიება

მარტივი ძიების ერთი კონკრეტული მაგალითია ცალკეული სიტყვაფორმის, მაგალითად: *"კაც"* ძიება კორპუსში. მარტივი ძიება, როგორც წესი, ემყარება გრაფემების თანმიმდევრულობის პრინციპს – კორპუსში იძებნება ყველა ის სიტყვაფორმა (ანუ ტოკენი), რომელიც შედგება კ, ა და ც გრაფემების თანმიმდევრული ჯაჭვისაგან. ასეთი კი შეიძლება იყოს სიტყვაფორმები: კაცი, კაცმა, კაცს..., საკაცე, საკაცემ, საკაცეს... კაცური, კაცურმა, კაცურს.... უკაცური, უკაცურმა, უკაცურს... კაცდება, დაკაცდა, დაკაცებულა... საკაცე, უკაცო, ვაუკაცი... და ა.შ. ანუ ყველა ის სიტყვაფორმა, რომელიც შეიცავს თანმიმდევრობას კ-ა-ც.

მარტივი ძიების მეორე მაგალითია მორფოლოგიური ელემენტის (მაგ.: სახელობითი ბრუნვის ნიშნის "ი" ~ Nom) ძიება კორპუსში. ასეთ შემთხვევაში მოიძებნება კორპუსში დადასტურებული ყველა სახელი, რომელიც სახელობით ბრუნვაში დგას. ამგვარი ძიების განხორციელება შესაძლებელია მხოლოდ იმ კორპუსებში, რომელიც მორფოლოგიური მეტაინფორმაციით არის აღჭურვილი ანუ სიტყვაფორმები ანოტირებულია ლინგვისტური მახასიათებლების მიხედვით. ამგვარი კორპუსები, როგორც წესი, მიემართება დამატებით რესურსებს: ლექსიკონებს, მორფემულ ლექსიკონთა ბაზას და მორფოლოგიური კომბინაციის წესების ბაზას.

მარტივი ძიების სხვა მაგალითია POS-ძიება (მაგ.: "N" ~ noun). ასეთი ძიების შედეგად იძებნება ყველა სიტყვა, რომელიც კლასიფიცირებულია, როგორც არსებითი სახელი. ამგვარი ძიების განხორციელება შესაძლებელია მხოლოდ იმ კორპუსებში, რომელიც შესაბამისი ლინგვისტური მეტაინფორმაციით „მეტყველების ნაწილი" (Part Of Speach) არის აღჭურვილი.

ბ) რთული ძიება

რთული ძიების ერთ-ერთ სახეს წარმოადგენს სინტაქსური სტრუქტურების, მაგ., სახელური ფრაზების (NP) სინტაქსური კვანძის მიხედვით ძიება (მაგ.: "NP" ~ noun phrase). ასეთ შემთხვევაში კორპუსში მოიძებნება ყველა ის კონტექსტი, რომელშიც სახელური ფრაზა დასტურდება. ამგვარი ძიების განხორციელება შესაძლებელია მხოლოდ იმ კორპუსებში, რომელსაც სინტაქსური ანალიზატორი (პარსერი) გააჩნია. რთული ძიების მექანიზმს საგანგებოდ განვიხილავთ ქვემოთ, ქართული ენის ეროვნული კორპუსის მაგალითზე.

სინტაქსური ძიების შემთხვევაში განსაკუთრებულ სიძნელეს ქმნიან ისეთი არარეპრეზენტაციული შემთხვევები, როგორიცაა ე.წ. გახლეჩილი სინტაგმები. მაგალითისათვის

შეგვიძლია მოვიტანოთ დიანა ანფიმიადის ლექსში *„მითოლოგემა“* გამოყენებული სინტაგმა **„გამოღმელთა სისხლი“**, რომელიც გახლეჩილი სახით არის წარმოდგენილი ლექსში:

„გამოღმელებმა შეისრუტეს ცრემლი გაღმელთა,

გაღმელებმა კი გამოღმელთა დალიეს სისხლი“

ლექსის ამ ფრაგმენტში პირველ სტრიქონში მოცემული სინტაგმა ***ცრემლი გაღმელთა*** პოსტპოზიციური სინტაგმაა, მეორე სტრიქონში მოცემული სინტაგმა ***გამოღმელთა სისხლი*** კი – პრეპოზიციულ სინტაგმას წარმოადგენს, რომელიც პრედიკატით – ზმნის გადმოცემული შემასმენლით არის გახლეჩილი.

თანამედროვე ქართულში სინტაგმის რამდენიმე სტრუქტურული მოდელი დასტურდება: მსაზღვრელი შეიძლება წინ უსწრებდეს ან მოსდევდეს საზღვრულს, მსაზღვრელი და საზღვრული შეიძლება გაიხლიჩოს წინადადების რომელიმე წევრით. ასეთ შემთხვევაში გვექნება სინტაგმის რეალიზების ორი სახე: ნორმატიული პრეპოზიციური სინტაგმა *vs.* გახლეჩილი პრეპოზიციური სინტაგმა, ნორმატიული პოსტპოზიციური სინტაგმა *vs.* გახლეჩილი პოსტპოზიციური სინტაგმა. კორპუსში სახელური ფრაზების ძიება კომპლექსური **სინტაქსური წესების ბაზისა** და მოდელების არსებობას მოითხოვს და მხოლოდ იმ კორპუსებში არის შესაძლებელი, რომელსაც სინტაქსური ანალიზატორი გააჩნია.

რთული ძიება სემანტიკური კონცეფციის კორპუსული გადამოწმების საშუალებასაც იძლევა, თუ იგი თავიდანვე ასახულია კორპუსის პრერეკეიზიტებში. მაგ.: *"მარილზე წასვლა"* ⇔ *"სიკვდილი"*, *"სულის გაფრთხობა"* ⇔ *"სიკვდილი"*, *"ფეხების გაფშეკა"* ⇔ *"სიკვდილი"*, *"გარდაცვალება"* ⇔ *"სიკვდილი"*, *"ჩაძაღლება"* ⇔ *"სიკვდილი"* და ა.შ. გარდა იმისა, რომ აქ მოყვანილი მაგალითები ერთ სემანტიკურ კონცეპტს „სიკვდილი“ მიემართებიან, ერთმანეთისაგან განსხვავდებიან სენტიმენტის თვალსაზრისით. კორპუსში „სიკვდილის“ სემანტიკური სინონიმების მოძებნა და კონტექსტში მათი სემანტიკური ანალიზი საშუალებას მოგვცემს დავაჯგუფოთ ისინი სენტიმენტის სკალაზე: სენტიმენტის თვალსაზრისით *"გარდაცვალება"* ნეიტრალურსაგნობრივად გადმოსცემს შესაბამის სემანტიკურ კონცეპტს. *"ფეხების გაფშეკა"* და *"სულის გაფრთხობა"* სემანტიკური კონცეპტის უარყოფითი კონოტაციით გადმომცემის შემთხვევაა, რომლის უკიდურესად მაღალ რეგისტრში ადის ფორმულირება *"ჩაძაღლება"*. *"მარილზე წასვლა"* კი უნდა განვიხილოთ, როგორც სემანტიკური კონცეპტის მეტაფორული გადმოცემის შემთხვევა. კორპუსში ამგვარი ანალიზის ჩატა-

რება საშუალებას მოგვცემს სენტიმენტანალიზატორის კონცეფცია შევიმუშაოთ, რათა მომავალში ავტომატურად განვახორციელოთ სენტიმენტანალიზი.

კორპუსის პრესტრუქტურირებიდან გამომდინარე, **რთული კორპუსული ძიება** შესაძლებელია მხოლოდ ანოტაციაში მითითებული ფენომენების მეშვეობით. ასეთი ფენომენები, როგორც წესი, კორპუსში „ფიზიკურად“ არის წარმოდგენილი. ამიტომ შეუძლებელია ისეთი გამოტოვებული ელემენტების საძიებო ცნებად ფორმულირება, როგორიცაა მაგ.:

ა) **ელიფსი** – *„ელენე ყავას სვამს, გიორგი ჩაის (~~სვამს~~)“*;

ბ) **ე.წ. Sluicing** ანუ გარიყული კითხვითი სიტყვა – *„ვილაც დავინახე, მაგრამ არ ვიცი ვინ (~~დავინახე~~)“*;

გ) **ნულოვანი მორფემები**, როგორიცაა მაგალითად **Ø**-ით მარკირებული მორფოსინტაქსური ელემენტები: მეორე სუბიექტური ან მესამე ობიექტური პირის ნიშნები ზმნაში: *მ-იხატავ* (Ø >S=შენ, *მ-* > IO=მე, Ø >DO=მას);

დ) **Ø-ით წარმოდგენილი სინტაქსური კონსტიტუენტები**, როგორიცაა მაგალითად pro drop-ელემენტები (გამოტოვებული ნაცვალსახელები): *„წავიდა და ალარც მოუკითხივარ“ < წავიდა (ის) და ალარც მოუკითხივარ (მას მე).*

რთული ძიების სპეციფიკურ შემთხვევას წარმოადგენს **გრამატიკალიზაციის** შედეგად წარმოქმნილი **ფუნქციური ელემენტების** ძიება კორპუსში. მაგალითისათვის შევადაროთ სიტყვაფორმის *„ერთი“* ფუნქციურ-სემანტიკური ანალიზი მომდევნო წინადადებებში:

1. *ერთი კაცი ამ საქმეს ვერ უშველის, ვერც ორი და ვერც ათი.*
2. *იყო ერთი კაცი.*
3. *ერთი გაბედოს და მოვიდეს.*

პირველ წინადადებაში – *ერთი კაცი ამ საქმეს ვერ უშველის, ვერც ორი და ვერც ათი* – სიტყვა *„ერთი“* რიცხვითი სახელია და რეალიზებულია, როგორც პრემოდიფიკატორი – ახდენს არსებითი სახელის *„კაცი“* დეტერმინირებას კვანტიტატიური თვალსაზრისით. მეორე წინადადებაში – *იყო ერთი კაცი* – სიტყვა *„ერთი“* განუსაზღვრელობითი ნაცვალსახელია, იგი შეიძლება ჩავანაცვლოთ განუსაზღვრელი ნაცვალსახელით *„ვინმე“*. ხოლო მესამე წინადადებაში – *ერთი გაბედოს და მოვიდეს* – იგივე *„ერთი“* მოდალური ფუნქციით გამოიყენება, მას ჩაენაცვლება მოდალური ნაწილაკი *„აბა“*. თუ *„ერთი“*-ს აქ აღნიშნული სამივე ფუნქცია თავიდანვე არ ჩაიდო კორპუსის პრერეკვიზიტებში, შეუძლებელი იქნება კორპუსში მისი დიფუნქციონირებული ძიება როგორც რიცხვითი სახელი, განუსაზღვრელი ნაცვალსახელი

ან მოდალური ელემენტი. ასეთ შემთხვევაში ან მხოლოდ მარტივი (ხაზობრივი) ძიებით უნდა დაეკმაყოფილდეთ და ძიების შედეგად გენერირებული კონკორდანსი ხელით დაამუშაოთ (ვინაიდან ასეთ შემთხვევაში კორპუსი შემოგვთავაზებს ყველა იმ კონტექსტს, რომელშიც ეს სიტყვაფორმა დასტურდება, ფუნქციის გათვალისწინების გარეშე) ან შეიმუშაოთ რთული სინტაქსურ-სემანტიკურ-პრაგმატიკული წესების კომპლექსი, რომელიც მოცემული ენობრივი ელემენტის (ტოკენის) ფუნქციური დიფერენციაციის საფუძველზე შეძლებს ჩვენ მიერ ფორმულირებული ამოცანის შესრულებას.

კორპუსული ძიების პრერეკვიზიტებს ქვემოთ საგანგებოდ განვიხილავთ ქართული ენის ეროვნული კორპუსის სპეციალურ ქვეთავში.

8. კორპუსლინგვისტიკის საკვანძო ცნებები

წინამდებარე თავებში განვიხილავთ კორპუსლინგვისტიკის ისეთ საკვანძო ცნებებს, როგორიცაა **ანოტაცია, ორაზროვნების მოხსნა (დისამბიგვირება)** და **მონაცემთა ვიზუალიზაცია.**

8.1. მონაცემის ახალი სახეობა – ანოტაცია

ანოტაცია **(annotation)** არის კორპუსის შემადგენელი კომპონენტი, რომელიც კორპუსული კვლევის განხორციელებისათვის აუცილებელ სამი სახის ინფორმაციას შეიცავს: **მეტამონაცემი (metadata), ტექსტის საგანგებო მარკირების ფორმა (textual markup)** და **ლინგვისტური ანოტაცია (linguistic annotation).** ამ უკანასკნელს (ლინგვისტური ანოტაცია) **ხაზთაშორისი გლოსირების პრინციპსაც** უწოდებენ. ვინაიდან მეტამონაცემებისა და ტექსტის მარკირების ფორმების შესახებ ზემოთ უკვე გვქონდა საუბარი, ამჯერად მხოლოდ ლინგვისტურ ანოტაციას შევეხებით.

ლინგვისტური ანოტაცია არის ანოტირების სახე – ენობრივი მონაცემების ლინგვისტური ანალიზის შედეგად მოპოვებული ინფორმაცია, რომელიც ანოტაციის ერთ ან რამდენიმე დონეზე (იხ. **ანოტაციის დონეები**) გადანაწილდება. ლინგვისტური ანოტაციის შემთხვევაში განარჩევენ დესკრიფციულ და ანალიტიკურ ანოტაციას. **დესკრიფციული ანოტაცია** ძირითადად ფორმალური ნიშნების მიხედვით ახდენს ენობრივი მონაცემების (ანუ სიტყვაფორმების) დახასიათებას – ბრუნვა, პირი, მწკრივი და ა.შ., **ანალიტიკური ანოტაციის** დროს კი ყურადღება ექცევა არაფორმალურ მახასიათებლებსაც, როგორიცაა მოდალობა, ევიდენციალობა, რეფერენციულობა. ლინგვისტური ანოტაცია ელექტრონული მონაცემის ლინგვისტური მეტაინფორმაცი-

ით გამდიდრების ერთ-ერთი ფორმაა და წარმოადგენს პირველად ენობრივ მონაცემებში ბუნებრივად კოდირებული ლინგვისტური ინფორმაციის **ხაზთაშორისი ვიზუალიზაციის** ყველაზე გავრცელებულ ფორმას.

კორპუსლინგვისტიკის განვითარების ერთ-ერთ უმნიშვნელოვანეს ეტაპად მიჩნეულია ლინგვისტური ანოტაციის პროცესის პროგრამული ავტომატიზაცია. ამ მიზნის მისაღწევად კორპუსის პროგრამულ ინფრასტრუქტურაში ჩაშენებულია სპეციალური ინსტრუმენტები, რომელთა საშუალებითაც ხორციელდება მიმართება სხვადასხვა სახის პირველად მონაცემთა ბაზებთან (როგორც წესი, ასეთია, მაგ., ლექსიკონი და მორფემათა ბანკი), რაც ლექსემათა ავტომატური ანოტირების საშუალებას იძლევა. ამგვარი მექანიზმით აღჭურვილ კორპუსს **ანოტირებული კორპუსი** ეწოდება.

ავტომატური ანალიზის საგანგებო პროგრამა – **მორფოლოგიური ანალიზატორი** – განახორციელებს სიტყვაფორმების მორფოლოგიურ ანალიზს მორფემათა ლექსიკონზე, მორფოსინტაქსური წესების ბანკსა და ლექსიკონზე დაყრდნობით.

ლექსიკონისა და მორფემათა ბანკის გარდა ავტომატური ანოტაციის დროს მოიხმარება ე.წ. **"წესების ბანკი"** (Rools), რომელიც წარმოადგენს მორფოლოგიური და მორფოსინტაქსური წესების ჩამონათვალს და ლოგიკურ მიმართებებს მათ შორის. მაგ.: სიტყვაფორმა "სპეციალობაზე", რომელიც ლექსიკონისა და მორფემული ბანკის მონაცემების მიხედვით ასე დაიშლება: სპეციალობა-ზე, (შეიცავს –ზე თანდებულს, ხოლო შესაბამისი ლექსემა ფუძის სახით "სპეციალობა" სახელობითი ბრუნვის ფორმად იდენტიფიცირდება), კვალიფიცირდება როგორც მიცემითი ბრუნვის ფორმა, ვინაიდან წესების ბანკის მიხედვით -ზე თანდებული სახელს მხოლოდ მიც. ბრუნვაში დაერთვის. წესების ბანკში ასევე აისახება კუმშვისა და კეცის წესები, მორფემათა სლოტები და მორფემათა თანმიმდევრობის კანონზომიერებები.

ორაზროვანი ფორმების გასარჩევად გამოიყენება მორფოსინტაქსური წესები. განვიხილოთ მაგ. ასეთი წინადადება: *„ის რაღაცას უპირებდა, თაეს არ ანებებდა"*.

სიტყვაფორმა ***ის*** ამ წინადადებაში (მორფოლოგიურ დონეზე) შეიძლება იყოს:

1. მესამე პირის ნაცვალსახელი ***ის*** სახელობით ბრუნვაში;
2. ჩვენებითი ნაცვალსახელი ***ის*** სახელობით ბრუნვაში.

ზემოთ მოყვანილ წინადადებას მორფოსინტაქსურ დონეზე თუ გავაანალიზებთ, ***ის*** კონსტიტუენტი კვალიფიცირდება, როგორც პირის ნაცვალსახელი, ვინაიდან წინადადებაში მის მომდევნოდ მდგარ კონსტიტუენტს *„რაღაცას"* კი არ ეთანხმება, არამედ წინადადების შემასმენელს *„უპირებდა"*: ***ის*** კონსტიტუენტი ჩვენებითი ნაცვალსახელი იქნებოდა, მაგალითად, ასეთ წინადადებაში: *„ის რაღაცა აჭუხებდა, თაეს არ ანებებდა"*, სადაც ***ის*** ბრუნვაში ეთანხმება კონსტიტუენტს

„რალაცა“ და, შესაბამისად, სახელური ფრაზის (*ის რალაცა)* ატრიბუტს წარმოადგენს. შდრ. ზემოთ მოყვანილი მაგალითის მოდიფიცირებული ვარიანტი: *„იმ რალაცის ჩვენებას არც აპირებდა, ფარ-ხმალს არ ყრიდა“.*

8.2. POS-ანოტაცია

დღესდღეობით ანოტაციის ყველაზე გავრცელებული ფორმა არის **POS-ანოტაცია (Part Of Speech Annotation)**. ზოგადად, ანოტაციის რამდენიმე სახეს განასხვავებენ:

ა) **ხელით შესრულებული (მანუალური) ანოტაცია** – სამეცნიერო ნაშრომებში გამოყენებული ტექსტის ლინგვისტური ანოტაცია, რომელიც, კორპუსული ანოტაციისაგან განსხვავებით, მანუალურად სრულდება და შრომატევადია.

ბ) **ნახევრად-ავტომატური ანოტაცია** – ძირითადად გამოიყენება კორპუსლინგვისტიკაში და მოითხოვს ავტომატური ანოტაციის შემდგომ ორაზროვნების მოხსნას.

გ) **ავტომატური ანოტაცია** – გამოიყენება კორპუსლინგვისტიკაში მორფოსინტაქსური ავტომატური ანალიზის (პარსინგის) დროს.

სამეცნიერო ლიტერატურაში ძირითადად გაერცელებულია **ხელით შესრულებული ლინგვისტური ანოტაცია**, რომლის მაგალითსაც ქვემოთ გთავაზობთ. ანოტაცია მოცემულ შემთხვევაში შედგება ოთხი დონისაგან: ა) ავთენტური ენობრივი დონე, სადაც წინადადება ავთენტური ფორმით არის მოცემული, ბ) ლინგვისტური ტრანსკრიფციის დონე, სადაც ლათინური შრიფტით არის გადაწერილი ავთენტური ენობრივი მონაცემი, გ) ლექსიკურ-გრამატიკული ანოტაციის დონე, სადაც ინგლისური თარგმანი და რამდენიმე გრამატიკული მახასიათებელია მოცემული და დ) წინადადების ინგლისური თარგმანი:

თხა-მ	ვენახ-ი	შეჭამ-ა
txa-m	*venax-i*	*šeč'am-a*
goat:ERG.Sg.	vineyard:NOM.Sg.	eat:AOR.S3

"The goat ate up the vineyard".

POS-ანოტაციისთვის გამოიყენება ხაზთაშორისი ანოტაცია, რომელიც რამდენიმე სახის შეიძლება იყოს: **ზოგადგრამატიკული ან მორფემულ-ხაზობრივი**. ზოგადგრამატიკული ხაზთაშორისი ანოტაციის შემთხვევაში მხოლოდ ძირითადი გრამატიკული მახასიათებლებია მოცემული, როგორც ეს ზემოთ მოყვანილ მაგალითში ჩანს. მორფემულ-ხაზობრივი ანოტაციის შემთხვევაში კი თითოული მორფემის ფუნქციურ-გრამატიკული მახასიათებელია წარმოდგენილი:

თხ-ებ-მა	ვენახ-ი	შე-უ-ჭამ-ეს
tx-eb-ma	*venax-i*	*še-u-čam-es*
goat: ST-Pl-:ERG.Sg.	vineyard.ST-NOM.Sg.	eat-Pr-OV-ST-S3Pl.(AOR)

"*The goats ate up somebody's vineyard*".

როგორც ეხედავთ, **ზოგადგრამატიკული ხაზთაშორისი ანოტაცია** ძირითადად ფორმალური ნიშნების მიხედვით ახდენს ენობრივი მონაცემების (სიტყვაფორმების) დახასიათებას, **მორფემულ-ხაზობრივი ანოტაციის** შემთხვევაში კი დროს ყურადღება ექცევა არაფორმალურ მახასიათებლებსაც, ამასთან სიტყვაფორმაში მორფემათა ანალიზი ემთხვევა მორფემათა თანმიმდევრობას. ეს უკანასკნელი განსაკუთრებით მნიშვნელოვანია ზმნების შემთხვევაში და ბევრად უფრო მოქნილია ქართულის, როგორც pro drop-ენის ლინგვისტური თვალსაზრისით ადეკვატურად აღსაწერად. შესაბამისად, **ზოგადგრამატიკული ხაზთაშორისი ანოტაცია** უნდა განვიხილოთ როგორც **ფორმალურ-დესკრიფციული ანოტაცია**, **მორფემულ-ხაზობრივი ანოტაცია** კი, როგორც **ანალიტიკური ანოტაცია**.

კორპუსში ლინგვისტური მონაცემების POS-ანოტაციის შემოტანამ და განსაკუთრებით ხაზთაშორისი ანოტაციის დამკვიდრებამ ამ მონაცემების ინტერნაციონალურ გამოყენებას ჩაუყარა საფუძველი და ფართოდ გაუღო კარი მეცნიერებს ტიპოლოგიური კვლევებისათვის მასალების მოსაძიებლად.

8.3. ანოტაციის დონეები

ლინგვისტური ანოტაცია შეიძლება ანოტაციის სხვადასხვა დონეზე გადანაწილდეს. **ანოტაციის დონეები (annotation levels)** არის **ლინგვისტური მეტაინფორმაციის** დამოუკიდებელ ბწკარებზე ასახვის ფორმა. ანოტაციის დონეები იმავდროულად წარმოადგენენ გრამატიკულად რელევანტურ დესკრიფციულ პერსპექტივებს მონაცემთა ბაზასთან მიმართებაში. ანოტაციის სრულყოფილი ფორმაა მრავალდონიანი ანოტაცია, რომელიც ლინგვისტური ინფორმაციის სხვადასხვა ენობრივ დონეზე გადანაწილების ეფექტურ ფორმას წარმოადგენს.

<u>მაგალითი 1:</u>

Lev1	თხა-მ	ვენახ-ი	შეჭამ-ა
Lev2	*txa-m*	*venax-i*	*šeč̣am-a*
Lev3	goat	vineyard	eat
Lev4	N	N	V-tr.
Lev5	ERG.Sg.	NOM.Sg.	AOR.S3
Lev7	A	DO	PRD
Lev10	"*The goat ate up the vineyard.*"		

მაგალითი 2:

Lev1	ეს	ვირუს-ის	მ-ა-ტარ-ებ-ელ-ი	კაც-ი
Lev2	*es*	*virus-is*	*m-a-ṭar-eb-el-i*	*ḳac-i*
Lev3	that	virus	carrying	man
Lev4	DPRON	N	PART.	N
Lev5	NOM.	GEN.	NOM.	NOM.
Lev6	ATT.	ATT.	ATT.	HEAD
Lev8	D1	-	-	

შდრ.:

Lev1	ამ	ვირუს-ის	მ-ა-ტარ-ებ-ელ-ი	კაც-ი
Lev2	*am*	*virus-is*	*m-a-ṭar-eb-el-i*	*ḳac-i*
Lev3	that	virus	carrying	man
Lev4	DPRON	N	PART.	N
Lev5	GEN.	GEN.	NOM.	NOM.
Lev6	ATT.	ATT.	ATT.	HEAD
Lev8	D1	-	-	

განმარტება: *Lev1 – ენობრივი მონაცემი ორიგინალის ენაში, Lev2 – ტრანსკრიფციის დონე, Lev3 – ლექსიკური დონე, Lev4 – POS-დონე (მეტყველების ნაწილები), Lev5 – მორფოლოგიური დონე, Lev6 – მორფოსინტაქსური დონე, Lev7 – სინტაქსური დონე, Lev8 – პრაგმატიკული დონე, Lev10 – თარგმანი ინგლისურად.*

მიუხედავად იმისა, რომ Lev6-ის მიხედვით ორივე ჩვენებითი ნაცვალსახელი "ეს" და "ამ" განსაზღვრებას წარმოადგენს და პრემოდიფიკატორია, პირველ მაგალითში "ეს" მიემართება სახელური სინტაგმის თავს (*"კაცი"*), მეორე მაგალითში კი "ამ" ნათესაობითში დასმული სახელის *"ვირუსი"* პრემოდიფიკაციას ახდენს და არა სახელური ფრაზის თავისა *"კაცი"*, რაც Lev5-ში მოცემული გრამატიკული მახასიათებლებით ადვილად "იშიფრება". ანუ ლინგვისტური ანოტაციის დონეებში მოცემული მახასიათებლები ვერტიკალურ და ჰორიზონტალურ ჭრილში ურთიერთგანმსაზღვრელი და ურთიერთგანმაპირობებელია.

დღეისათვის არსებული კორპუსები ერთმანეთისაგან, როგორც წესი, განსხვავდებიან როგორც ანოტაციის დონეების რაოდენობის, ისე დონეების იერარქიული სტრუქტურის თვალსაზრისითაც.

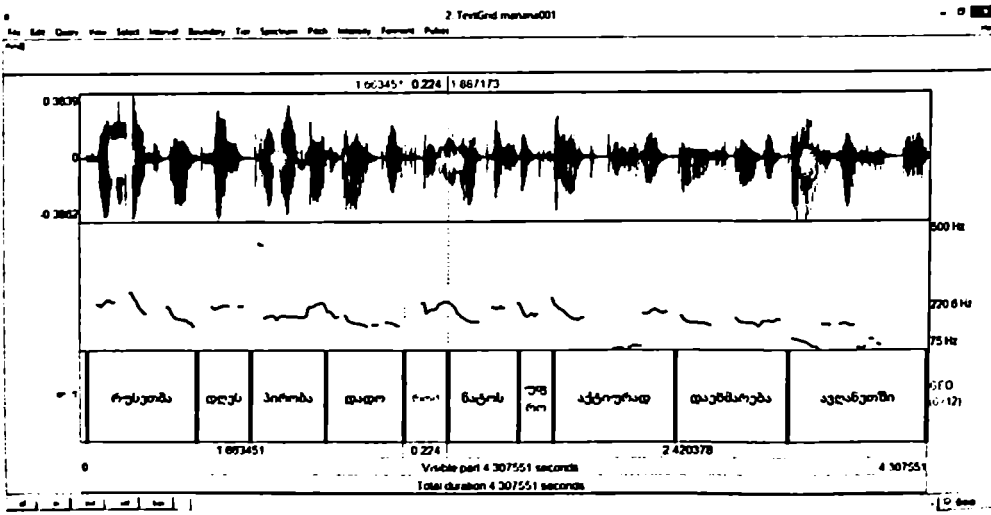
კორპუსის მოხმარების ეფიციენტურობა ლინგვისტური კვლევების თვალსაზრისით ლინგვისტური ანოტაციის ხარისხზე არის დამოკიდებული. ქვემოთ ჩვენ შევეცდებით ვაჩვენოთ, თუ რამდენად მნიშვნელოვანია კორპუსის აღჭურვა მრავალდონიანი ანოტირების სისტემით.

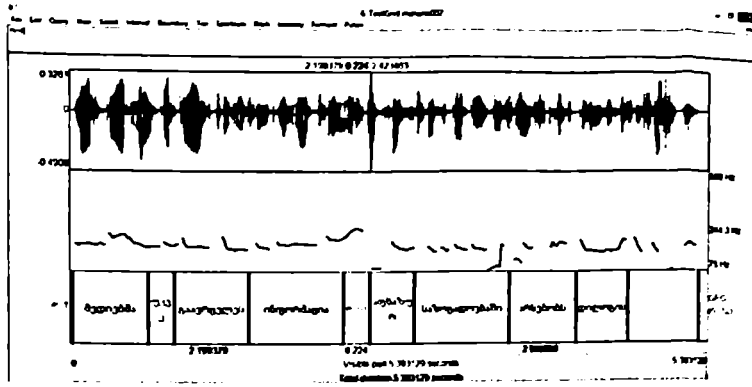
მაგალითისათვის განვიხილოთ რამდენიმე წინადადება:

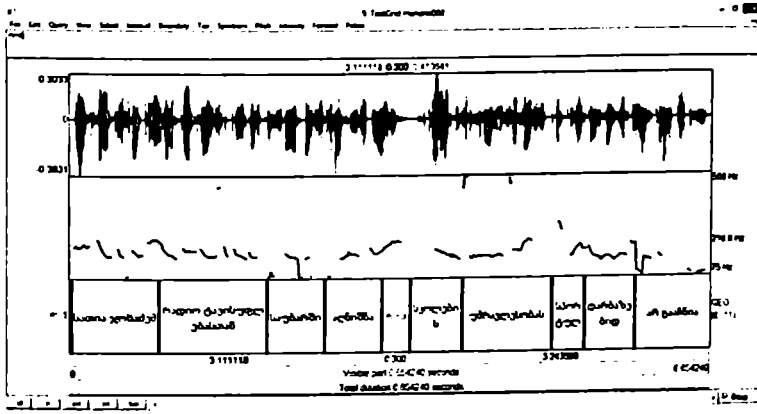
1. *რუსეთმა დღეს პირობა დადო, რომ ნატოს უფრო აქტიურად დაეხმარება ავღანეთში.*
2. *მედიებმა უკვე გააერცელეს ინფორმაცია, რომ აფხაზურ საზოგადოებაში არსებობს დიალოგის მზაობა.*
3. *ნათია ჯოხაძემ რადიო თავისუფლებასთან საუბარში აღნიშნა, რომ სკოლების უმრავლესობას სპორტული დარბაზები არ გააჩნია.*
4. *გუშინ რუსიკომ დარეკა და დაიბარა, რომ დაბრუნდება, აუცილებლად შემეხმიანოსო.*

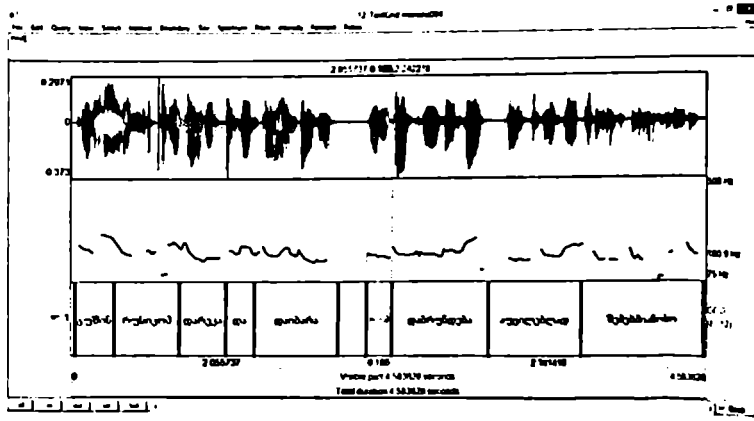
ამ წინადადებებს რამდენიმე საერთო ნიშანი გააჩნიათ: ოთხივე წინადადება რთული ქვეწყობილი წინადადებაა, ოთხივეში დასტურდება ელემენტი *რომ*, რომელიც დაქვემდებარებული წინადადების თავში დგას. ჩვენ რამდენჯერმე ჩავატარეთ ცდა სხვადასხვა ასაკისა და პროფესიის ადამიანებთან – ჩაიწერეთ მათ მიერ წაკითხული ზემოთ მოყვანილი წინადადებები და კომპიუტერულად დავამუშავეთ ისინი სუპრასეგმენტური ანალიზის თვალსაზრისით. საინტერესოა, რომ სუპრასეგმენტური ანალიზის შედეგები დაემთხვა ჩვენს მოსაზრებას *რომ* ელემენტის ფუნქციური სემანტიკის თაობაზე: პირველ სამ წინადადებაში *რომ* მაქვემდებარებელი კავშირია, მეოთხეში კი – დროის გარემოების ფუნქციით გამოიყენება, რასაც სუბსტიტუციის მეთოდის საშუალებით ადვილად დავასაბუთებთ: *გუშინ რუსიკომ დარეკა და დაიბარა, როდესაც დაბრუნდება, აუცილებლად შემეხმიანოსო.*

შდრ.:









როგორც სუპრასეგმენტური ანალიზის შედეგები გვიჩვენებს, მეოთხე წინადადება საგრძნობლად განსხვავდება წინა სამისაგან. ლინგვისტური ანალიზის მიხედვით პირველ სამ წინადადებაში *რომ* მაქვემდებარებელი კავშირია, მეოთხე წინადადებაში კი – დროის გარემოებას წარმოადგენს. „*რომ*“ ამ შემთხვევაში „*როცა*“ ან „*როდესაც*“ – ფუნქციური ეკვივალენტია, თუმცა წინადადებაში ტოპოლოგიური თვალსაზრისით, მანქანური დამუშავების დროს, არაფრით განსხვავდება წინა სამ წინადადადებაში გამოყენებული მაქვემდებარებელი კავშირისაგან. სუპრასეგმენტური ანალიზის დონის შემოტანა „წაკითხვადს“ გახდის მანქანურ დონეზე ამ განსხვავებას, გაცილებით დახვეწს ძიების პროცესს (ანუ ავტომატურად გამოავლენს „*რომ*“-ის ხმარების ისეთ შემთხვევებს, როგორიც მეოთხე მაგალითში გვაქვს აღწერილი) და საგრძნობლად გაზრდის კორპუსის ადეკვატურად გამოყენების ეფექტიანობას. შესაბამისად, ანოტაციის დონეები სასურველია შემდეგნაირად განიერცოს:

1 – ენობრივი მონაცემი ორიგინალ ენაში

2 – ტრანსკრიფციის დონე

3 – ლექსიკური დონე (თარგმანი)

4 – POS-დონე (მეტყველების ნაწილები)

5 – მორფოლოგიური დონე

6 – მორფოსინტაქსური დონე

7 – სინტაქსური დონე

8 – პრაგმატიკული დონე

9 – სუპრასეგმენტური დონე

10 – სემანტიკური დონე

ანოტაციის დონეების გაზრდა მნიშვნელოვანია ქართულში ნულით გადმოცემული წინადადების წევრების ან ელიფსური კონსტრუქციების კვლევის პერსპექტივიდანაც.

ქართული ენა pro drop-ენათა რიგს მიეკუთვნება, რაც იმას ნიშნავს, რომ ნაცვალსახელით გადმოცემული სუბიექტი – ქართულის შემთხვევაში კი არა მხოლოდ სუბიექტი, არამედ პირდაპირი და ირიბი ობიექტიც – ელიმინირებას ექვემდებარება ანუ წინადადების ეს წევრები შეგვიძლია გავანულოთ. თუმცა, არის ისეთი შემთხვევები, როდესაც მათი ელიმინირება შეუძლებელია – წინადადებაში მათი პოვნიერება აუცილებლობით არის განპირობებული. თუ რაში გამოიხატება ეს აუცილებლობა (როდის არის შესაძლებელი pro drop-ელემენტების განულება), ქართულისათვის არ

არის გამოკვლეული. როგორ მოვიპოვოთ კორპუსში ემპირიული მასალა განულებული რეფერენტების შეთხვევაში?

სუბიექტის (resp. ობიექტის) პოვნიერებას pro drop-ენებში რეფერენცია უზრუნველყოფს. მოკლედ დავახასიათებთ რეფერენციულობის გამოხატვის ზოგიერთ საშუალებას ქართულში.

რეფერენცია ქართულში, განულებული სუბიექტის (resp. ობიექტის) შემთხვევაში, ზმნის პოლიპერსონალიზმის ხარჯზე გამოიხატება: სუბიექტის და ობიექტის პირის ნიშნები (ან მათი კომბინაცია) მიუთითებენ შესაბამისი აქტანტის პოვნიერებაზე. განვიხილოთ ორი მაგალითი: *„სულ დაიბრა“* და *„რა არ ვუთხარი, მაგრამ ვერ დავაჯერე“*. ორივე წინადადებაში გამოტოვებულია წინადადების ძირითადი წევრები – პირველში სუბიექტი და ირიბი ობიექტი, მეორეში კი სუბიექტი და პირდაპირი ობიექტი. მიუხედავად ამისა, მორფოსინტაქსურ და სემანტიკურ-პრაგმატულ დონეზე ადვილად აღდგება ყველა:

1. [ის] სულ დაიბნ-ა
2. [მე] რა არ ვ-უ-თხარი [მას], მაგრამ [მე] [ის] ვერ და-ვ-ა-ჯერე.

რეფერენციულობას ამ შემთხვევაში უზრუნველყოფენ მორფოსინტაქსური ელემენტები, როგორიცაა პირის (სუბიექტური პირის ნიშანი ზმნაში: *დაიბნ-ა, და-ვ-ა-ჯერე*) ან ქცევის ნიშანი (ქცევის ნიშანი ზმნაში: *ვ-უ-თხარი, და-ვ-ა-ჯერე*). თუმცა, ამ უკანასკნელში პირდაპირი ობიექტი ზმნაში არ არის აღნიშნული, ნული მორფემით არის გადმოცემული. ანოტაციის სისტემაში ნულით წარმოდგენილი *pro* ელემენტების ასახვა მსგავს მაგალითებში არსებით მნიშვნელობას იძენს და კორპუსში ენის სხვადასხვა დონეზე კვლევის შედეგიანობის ხარისხს ბევრად ამაღლებს.

რეფერენციულობის გამოხატვის მორფოსინტაქსური საშუალებები (პირისა და ქცევის ნიშნები) ქართულისათვის ეფექტურად შეგვიძლია გამოვიყენოთ, მაგრამ არა ყოველთვის. განვიხილოთ ერთი მაგალითი ბესო ხვედელიძის მოთხრობიდან "ქალალდის თვითმფრინავები":

ნელა მოატრიალეს მისკენ სახე, კაი ხანს აკვირდებოდნენ.

ერთი შეხედვით, ამ წინადადებაში გამოტოვებულია სუბიექტი, რომლის პოვნიერებასაც ზმნური ფორმები *მოატრიალ-ეს* და *აკვირდებოდ-ნენ* გვიდასტურებს. ორივე შემთხვევაში ზმნაში გამოიყოფა მესამე სუბიექტური პირის მრავლობითი რიცხვის სუფიქსი – *-ეს* და *-ნენ*. მიუხედავად ამისა, ზმნის სუბიექტი ორივე შემთხვევაში ერთი პიროვნებაა, რაზედაც შესაბამისი კონტექსტი მიგვითითებს:

"კინოთეატრის ფოიეში ხმაურით შეპაკუნდა. ბილეთის აღებაზე არც უფიქრია, ისე შეეარდა დარბაზში. დიდის სეანსი გადიოდა. გეზი ბოლო რიგისკენ აიღო. მიუნეელ სიბნელეს ხელისცეცებით იკვალაედა. ბოლოში კედელს გადაეყუდა, განაპირა სკამზე დაეშვა და ეკრანს მზერა გაუსწორა.

„მთავარს თვალებით ვერ დაინახავ!" – ჩაუკრა თვალი ეკრანიდან ფილმის გმირმა.

უცებ გათბა. თითქოს აქამდე უცხო სინათლით აივსო. მერე გვერდით მჯდომისკენ გააპარა თვალი და გაიღიმა.

ნელა მოატრიალეს მისკენ სახე. კაი ხანს აკვირდებოდნენ.

– მოეედი. – მოჭრით თქვა ამან და ტუჩები კეკლუცად მოპრუწა.

არაფერი უპასუხეს. მხოლოდ თაეი დააქნიეს მძიმედ და კელაე ეკრანისკენ მიაბრუნეს. ახლაღა შეამჩნია, დარბაზში მარტონი ისხდნენ."

წინადადებაში ე.წ. უპირო ზმნის კონსტრუქცია გვაქვს – *მოატრიალ-ეს* და *აკვირდებოდ-ნენ* ფორმები რეფერირებას ახდენენ არა სუბიექტის კონკრეტულ სიმრავლეზე (**ბევრმა იქ მყოფმა** მოატრიალა თავი და **ბევრი იქ მყოფი** აკვირდებოდა), არამედ ერთ კონკრეტულ სუბიექტზე, რომელიც უცხო რჩება მკითხველისთვის, ვინაიდან ავტორს ეს პერსონაჟი ჯერ არ შემოუყვანია – მას არც სახელი აქვს, არც სქესი – მისი პიროვნება უცნობია.

ზმნის მრავლობითი რიცხვის ფორმის გამოყენება არაპირდაპირი მნიშვნელობით სხვა შემთხვევებშიც დასტურდება. ამგვარი ზმნური ფორმები – ზმნა მესამე პირის მრავლობით რიცხვში – ხშირად მოდალობასაც ითავსებს. მაგ., მრავლობითი რიცხვის ფორმა მომდევნო წინადადებებში მთქმელის სუბიექტურ ინტენციას – მიზანშეწონილობას ან აუცილებლობას გამოხატავს:

მაგალითი 1:

ყურძენს სექტემბრის ბოლოს კრეფენ = მიზანშეწონილია, რომ ყურძენი სექტემბრის ბოლოს დაიკრიფოს და არა ყურძენს [ისინი ვ3მრ.] სექტემბრის ბოლოს კრეფენ.

ამგვარ წინადადებაში სუბიექტი არ აღდგება, ვინაიდან იგი არც იგულისხმება – ზმნა უპიროა (შდრ. შესაბამისი კონსტრუქცია გერმანულში: Man sammelt die Weintraube Ende Oktober).

მაგალითი 2:

ასეთ რეპლიკას უპასუხოდ არ ტოვებენ = ასეთ რეპლიკას, როგორც წესი, უპასუხოდ არ ტოვებენ / ასეთ რეპლიკას პასუხი უნდა გაეცეს და არა *ასეთ რეპლიკას უპასუხოდ არ ტოვებენ [ისინი S3მრ.].*

კორპუსლინგვისტიკის განსაზღვრებისას ზემოთ უკვე აღვნიშნეთ, რომ კორპუსლინგვისტიკა ხშირად კვლევის სრულიად ახალ პერსპექტივებს ხსნის, რომელიც დიდი მოცულობის მონაცემების გადამუშავებას მოითხოვს, რაც მხოლოდ ანოტაციის უფრო რთული სტრუქტურის შემთხვევაშია შესაძლებელი. ამ და სხვა მსგავსი მაგალითების განხილვა, თვალნათლივ გვიჩვენებს, თუ რამდენად მნიშვნელოვანია ნულით წარმოდგენილი აქტანტების ანოტაციაში ასახვა, რათა ნულის სინტაქსის პრობლემა ქართულში კორპუსზე დაყრდნობით ვიკვლიოთ.

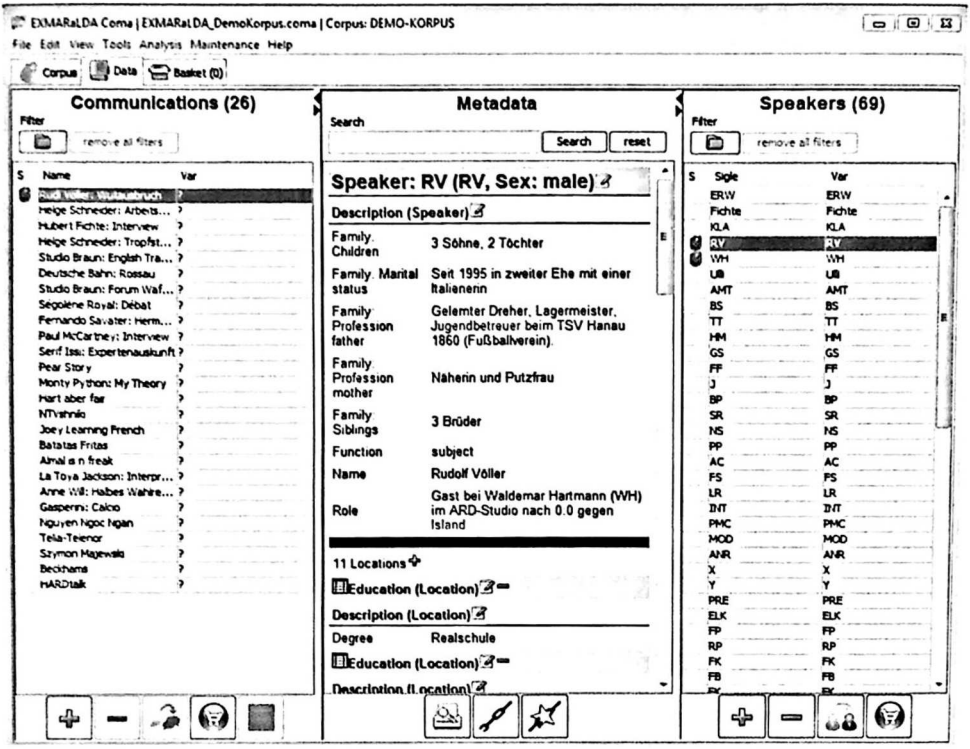
საგულისხმოა, რომ მოდალობის ფუნქციით გამოყენებული ზმნური ფორმები უღლილების ძალიან შეზღუდულ პარადიგმას გვიჩვენებენ: ისინი, როგორც წესი, არ იცვლებიან პირში და რიცხვში, ასევე შეზღუდულად ან საერთოდ არ გააჩნიათ სხვა მწკრივის ფორმები.

ლინგვისტური ანოტაციის დონეების სისტემის შექმნისას ანოტაციის დონეებში სასურველია დამატებითი დონის შემოტანა უცხო სიტყვების აღსანიშნად, რათა ისინი ძიების დროს ავტომატურად გამოეყოს ავთენტურ ენობრივ მასალას. ეს საკითხი ერთ-ერთ მნიშვნელოვან პუნქტს წარმოადგენს კორპუსებში ანოტაციის სისტემის შემუშავების დროს, განსაკუთრებით ისეთი კორპუსებისათვის, რომლებიც ბილინგვურ ტექსტებს შეიცავენ.

ასევე მნიშვნელოვანია ტექსტებში შემოკლებების მარკირება. როგორც წესი, ასეთ შემთხვევაში დგება შემოკლებების სია (სახელების, მაგ.: ალ.<ალექსანდრე, ან გვარების, მაგ.: ჭავჭ.< ჭავჭავაძე, ან დღეების: ოთხ. < ოთხშაბათი და ა.შ.). ანოტაციის ერთ-ერთ გავრცელებულ სისტემას წარმოადგენს ტექსტის მარკირების საგანგებო ნიშნულების გაფართოებული სისტემა დისკურსის ანოტაციისათვის EXMARaLDA (Extensible Markup Language for Discourse Annotation)[118]. EXMARaLDA თავდაპირველად ჰამბურგის უნივერსიტეტში მიმდინარე ერთი სამეცნიერო პროექტის[119] ქვეპროექტის ფარგლებში განვითარდა და მიზნად ისახავდა მულტილინგვალური მონაცემების ელექტრონული დამუშავებისა და ანალიზის მეთოდების შემუშავებას. ტესტის მარკირების აღნიშნული სისტემა წარმატებით გამოიყენება სხვა კორპუსებისათვისაც.

[118] დაწვრილებით იხ. http://www.exmaralda.org

[119] "Mehrsprachigkeit" (SFB 538).



EXMARaLDA – კორპუსის მენეჯერი

ანოტაციის აღნიშნული სისტემა საგანგებოდ შეიქმნა ზეპირმეტყველების მასალების კორპუსული დამუშავების მიზნით და შედგება მონაცემთა საგანგებო ფორმატის, კონცეპტების და საგანგებო ტექნოლოგიური ინსტრუმენტებისაგან, რომელიც საშუალებას იძლევა ზეპირი მეტყველების ნიმუშები ავტომატურად დავამუშაოთ ტრანსკრიფციისა და ანოტაციის თვალსაზრისით. EXMARaLDA-ს ტრასკრიფციისა და ანოტაციის სისტემა დღეისათვის თავისუფალ რესურსს წარმოადგენს და 2011 წლამდე მუდმივად ვითარდებოდა და იხვეწებოდა ჰამბურგის ენობრივი კორპუსების ცენტრში (Hamburger Zentrum für Sprachkorpora), 2011 წლის ივლისიდან კი მის განვითარებაზე ზრუნავს მანჰაიმის გერმანული ზეპირმეტყველების ინსტიტუტი (Institut für Deutsche Sprache in Mannheim).

8.4. დისამბიგვირება – ორაზროვნების მოხსნა

ცნება დისამბიგვირება ლათინური სიტყვაა და ქართულად ორაზროვნების მოხსნას ნიშნავს. საქართველოში კორპუსლინგვისტიკის განვითარების პირველსავე

დღიდან დაიწყო ძიება ამ ცნების "გასაქართულებლად". დღეისათვის შემოთავაზებული ქართული ვერსიებიდან (გაერთმნიშვნელოვნება, მონოსემიზაცია, ომონიმიის მოხსნა) ყველაზე მისაღებ ქართულ ეკვივალენტად მიგვაჩნია "ომონიმიის მოხსნა", თუმცა ვერც ეს უკანასკნელი ვერ ასახავს ზუსტად დისამბიგვირების, როგორც კორპუსლინგვისტიკაში ერთ-ერთი უმნიშვნელოვანესი პროცესის, ფუნქციურ არსს.

დისამბიგვირება არ არის კორპუსლინგვისტიკის სპეციალური ცნება. ამ ფენომენს ტრადიციული ლინგვისტიკა უკვე დიდი ხანია იცნობს და არაერთი თეორიული ნაშრომიც არის შექმნილი მის შესახებ. ლექსიკოლოგიაში ლექსიკური დისამბიგვირება სალექსიკონო ერთეულთა სტრუქტურირების ერთ-ერთ გადამწყვეტ ფაქტორად არის მიჩნეული. სემიოტიკაში კი იგი ნიშანთა და სიმბოლოთა სისტემატიზაციის ერთ-ერთ არსებით პრობლემას წარმოადგენს.

დისამბიგვირება განსაკუთრებით რთული განსახორციელებელი აღმოჩნდა ენის ტექნოლოგიების – ენის ავტომატური ანალიზისა და სინთეზის, სპელჩეკერის, ავტომატური თარგმანისა და ტექნოლოგიური სახის სხვა პროდუქტების შექმნის პროცესში – დისამბიგვირების ავტომატიზაცია ლამის კორპუსლინგვისტიკის აქილევსის ქუსლად იქცა სპეციალისტებისათვის.

საილუსტრაციოდ განვიხილოთ რამდენიმე წინადადება.

შდრ. მაგალითები სახელური მორფოლოგიიდან:

ა) *პროფესორ ვახტანგ იმნაიშვილს დაბადების 70 წელი შეუსრულდა.*

ბ) *პროფესორ ვახტანგ იმნაიშვილის სახელი უცხოეთის სამეცნიერო წრეებში კარგადაა ცნობილი.*

გ) *პროფესორ ვახტანგ იმნაიშვილით ამაყობს დღეს ქართველოლოგია.*

დ) *იგი ჩვენთვის ყოველთვის პროფესორ ვახტანგ იმნაიშვილად დარჩება.*

ოთხივე წინადადებაში ლექსემა „პროფესორი" გაუფორმებელი სახით არის წარმოდგენილი, მიუხედავად იმისა, რომ იგი სხვადასხვა ბრუნვაში (მიცემითში, ნათესაობითში, მოქმედებითში და ვითარებითში) დასმულ სახელს ერთვის, როგორც მსაზღვრელი.

შდრ. მაგალითები ზმნური მორფოლოგიიდან:

ზმნური ფორმა "*გაქებთ*" ოთხი მორფოსინტაქსური კონსტრუქციის ამსახველია:

ა) *გაქებთ* მე თქვენ (S1 O2Pl)

ბ) *გაქებთ* ის თქვენ (S3 O2Pl)

გ) *გაქებთ* ჩვენ თქვენ (S1Pl O2Pl)

დ) *გაქებთ* ჩვენ შენ (S3Pl O2)

თუ ქართული ენის, როგორც pro-drop-ენის შესაძლებლობას გავითვალისწინებთ, ნაცვალსახელით გამოხატული სუბიექტი და ობიექტი გაანულოს, მაშინ არც სინტაქსურ დონეზე არ გაგვაჩნია არანაირი „ნიშანი“, რომელიც *გაქცებთ* ფორმის სწორი ანალიზის ავტომატიზაციის საშუალებას მოგვცემს წინადადების ფარგლებში.

აქედან გამომდინარე, ალბათ, უფრო ზუსტი იქნებოდა დისამბიგვირებისათვის **მორფოსინტაქსური ომონიმიის მოხსნა** გვეწოდებინა. თუმცა ამ ტერმინის სიზუსტეც ეჭვქვეშ დადგება, თუ გავითვალისწინებთ დისამბიგვირების ისეთ შემთხვევებს, როგორიცაა მორფოლოგიური საშუალებების გამოყენება ლექსიკური ომონიმიის შემთხვევაში. აქ ჩვენ მხედველობაში გვაქვს ისეთი ომონიმური ფორმები, როგორიცაა "*თვალი*", როგორც მხედველობის ორგანო (მაგ. ადამიანის *თვალი*) და "*თვალი*", როგორც სატრანსპორტო საშუალების ერთ-ერთი კომპონენტი (მაგ. ველოსიპედის *თვალი*). ლექსიკური ომონიმიის მოსახსნელად ქართული ენა მორფოლოგიური პარადიგმის აგების ისეთ შესაძლებლობებს იყენებს, როგორიცაა კუმშვა: "*თვალი*" პირველი მნიშვნელობით არ იკუმშება, ხოლო მეორე მნიშვნელობით – იკუმშება. ისეთ ბრუნებში, სადაც კუმშვა არ ხორციელდება (სახელობითი, მოთხრობითი, მიცემითი...), დისამბიგვირება აუცილებელია, ოღონდ ამ შემთხვევაში ეს იქნება არა მორფოლოგიური, არამედ **ლექსიკური ომონიმიის მოხსნა**.

მსგავსი მაგალითები სხვაც მოგვეპოვება ქართულში, მაგალითად, **კვარტალი**. ქართული ენის ლექსიკონში ეს ლექსემა შეტანილია როგორც არაკუმშვადი სახელი და განმარტებულია როგორც *1. ქალაქის ნაწილი, რომელიც მოქცეულია ოთხი ურთიერთ-გადამკვეთი ქუჩის შიგნით; 2. წლის მეოთხედი, სამი თვე.*[120]

ლექსიკური ომონომიის დასაძლევად ქართული ენა გარკვეულ შემთხვევებში ბრუნების ტიპებს იყენებს: პირველი მნიშვნელობით სიტყვა *კვარტალი* უკუმშველია, მეორე მნიშვნელობით კი – კუმშვადი, რასაც ქართული ენის ეროვნული კორპუსის მასალებიც გვიდასტურებს. ვინაიდან დისამბიგვირების რამდენიმე სახე არსებობს, ამ ტერმინის ნაცვლად შემოგვაქვს ტერმინი **ომონიმიის მოხსნა,** რომელსაც საჭიროების შემთხვევაში დაემატება მსაზღვრელი: მორფოლოგიური ომონიმიის მოხსნა, მორფოსინტაქსური ომონიმიის მოხსნა, ლექსიკური ომონიმია და ა.შ.

ომონიმიის მოხსნის გაცილებით რთული შემთხვევა გვაქვს მაშინ, როდესაც სიტყვას ერთდროულად გააჩნია როგორც ლექსიკური, ისე ფუნქციური მნიშვნელობა., როგორც, მაგალითად, *ერთის* შემთხვევაში. იგი წინადადებაში შეიძლება გამოყენე-

[120] ციტატა მოტანილია ქართული ენის განმარტებითი ლექსიკონის რვატომეულის ელექტრონული ვერსიიდან. იხ. http://www.ena.ge/

ბული იყოს როგორც რიცხვითი სახელი, განუსაზღვრელობითი სახელი, მოდალური ელემენტი და როგორც ქვემოთ მოყვანილი მაგალითები გვიჩვენებს, მისი გამოყენების სხვა შემთხვევებიც დასტურდება. ასეთ შემთხვევაში ომონიმიის მოხსნა გულისხმობს არა რომელიმე კონკრეტული, მაგალითად, მორფოლოგიური ან ლექსიკური ომონიმიის მოხსნას, არამედ გაცილებით კომპლექსურია და ლექსიკური და ფუნქციურ-სემანტიკური ორაზროვნების განეიტრალებას, ანუ მოცემული სიტყვაფორმის სემანტიკურ და ფუნქციურ დონეზე ანალიზს მოითხოვს. შდრ.:

(1) ისაა ადგომას ვაპირებთ, გავიგოთ ერთი, რა მოხდაო – მოაკაკუნა ეიღაცამ.

(2) ფორიად გავაცანი სეთ'ურს დათა. შეგვიპატიჟა. იქ თაბაგარი დაგეხვდა ერთი. – ჩემი მარჯვენა ხელიაო.

(3) ერთი კია, ს'უფრას ქორიეით დასტრიალებდა, ჩიტის რძეს არ გეაკლებდა.

(4) მე აქეთ-იქით ვიყურებოდი, სარკეს ვეძებდი, ერთი ჩამახედა, ნეტავი, რას ეგავარ ისეთს, რომ ეს მახირჯი დედაბერიც კი შეეწუხე და შეეშარე-მეთქი.

(5) რეა ჯერდა წამომეყვანა, მისთვის ერთი დამეტოვებინა და არაფერს იტყოდა მაშინ.

(6) რას მოეცსწარი ამას, დასწყეელოს ზეცა-ჯფალმა! სალდათობა რომ არ მინდოდა, გააბრაგებამ მიტომ მომიწია... შეხედე. ერთი.

(7) ერთი გამაგებინა, ვახტანგა შალითურო, ეგრე მწარედ რამ დაგბოლმა და მერე გინდ მომკლა!

(8) ჯერდა დავაკვირდე, ერთი, ამას.

(9) ერთი თხუთმეტ-თექვსმეტი წლის იქნებოდა..

(10) დამიხედეთ, ერთი, ამ ნაძირალებს, ამათ?! გაეთრიეთ აქედან!

(11) ცოტა ხანი კიდევ იტრიალეს, ერთი კარგა იქვიფეს და წავიდნენ.

(12) კოსტა, ერთი აქეთ მობრძანდი, საქმე მაქვს!

(13) ერთი კი შეერხა წარბი დასტურიძეს და მერე ისე გაიკვირვა ჩემი ნათქვამი, რომ ეჭვი შემეპარა, იქნებ ეს კაცი დასტურიძე არ არის, ანდა დასტურიძეა და ამ საქმისა მართლაც არა იცის რა-მეთქი.

(14) ერთი ხანი კიდევ უცქირა და ახლა მე შემომხედა.

თუ რა ფუნქციური სემანტიკის მატარებელია მოცემულ წინადადებებში სიტყვა „ერთი“, საგანგებო კელევის საგანია და აქ მასზე აღარ შევჩერდებით. თუმცა, თვალსაჩინოებისათვის გვინდა მოვიყვანოთ ზემოთ მოყვანილი რამდენიმე წინადადების ავტომატური ანალიზის მაგალითი ქართული ენის ეროვნული კორპუსიდან, სადაც დაძლეულია დისამბიგვირების პრობლემა.


ქეგკ
ქართული ეროვნული კორპუსი
The Georgian National Corpus
გაანალიზება


ქეგკ
ქართული ეროვნული კორპუსი
The Georgian National Corpus
გაანალიზება


ქეგკ
ქართული ეროვნული კორპუსი
The Georgian National Corpus
გაანალიზება

პირველ შემთხვევაში – მაგალითი (9) – *ერთი* ფუნქციურ-სემანტიკურად აპროქსიმატიული ზმნისართია და ზმნისართებს *დაახლოებით/თითქმის* ჩაენაცვლება, მეორე შემთხვევაში – მაგალითი (14) – ზედსართავი სახელის *ცოტა* ფუნქციური ეკვივალენტია, მესამე შემთხვევაში კი – მაგალითი (7) – მოდალური (ნატერითი) ნაწილაკის *ჩუტაჟ* ფუნქციას ითავსებს.

აქვე გვინდა აღვნიშნოთ, რომ ამგვარი შემთხვევები არც თუ იშვიათად გვხვდება ქართული ენის კორპუსში და ომონიმიის მოხსნის პრობლემის გადაჭრა საგრძნობლად დახვეწს ქართული ენის ეროვნულ კორპუსში სინტაქსური ანალიზის პროგრამას.

8.5. მონაცემთა ვიზუალიზაცია

ტრადიციული ლინგვისტიკისაგან განსხვავებით კორპუსლინგვისტიკა არა მხოლოდ ფაქტების, ანუ ენობრივი მონაცემების აღწერით არის დაკავებული, არამედ მონაცემებს შორის არსებულ მიმართებებსაც აღწერს და რაც მთავარია, შეუძლია ამ მიმართებების ვიზუალიზაცია.

რა არის ვიზუალიზაცია?

ვიზუალიზაცია (visualisation) არის აბსტრაქტული მიმართებების ოპტიკური საშუალებებით ასახვა. ვიზუალიზაციის მეთოდი წარმატებით გამოიყენება დებულებების ან შედეგების თვალსაჩინოდ წარმოდგენის მიზნით მეცნიერებაში, მედიებში, საყოფაცხოვრებო სფეროში. დღეს უკვე აღარ დაობენ ინფორმაციის ვიზუალიზაციის მნიშვნელობის, კერძოდ, მისი გაელენის შესახებ საზოგადოების ცნობიერების ჩამოყალიბებაზე ან ცვლილებაზე. არც ის ფაქტი არ არის სადაო, რომ მედიასაშუალებები საზოგადოებრივი ცნობიერების მანიპულირების ერთ-ერთ უძლაერეს იარაღს წარმოადგენს.

არსებობს ვიზუალიზაციის მეთოდები და ინსტრუმენტები, რომლებიც არსებითად განსხვავდებიან ერთმანეთისაგან მიზნის, სტრატეგიისა და გამოყენების სფეროების მიხედვით, თუმცა საერთოა ვიზუალიზაციის ზოგადი პრინციპი: იგი **ცხადად და სხარტად უნდა გადმოსცემდეს ფუნქციურ მიმართებებს რელევანტურ ფაქტებს ან მოვლენებს შორის.** ვიზუალიზაციის უძველეს ფორმას მიეკუთვნება პიქტოგრამა, ხოლო უახლესს – ლოგო, დიაგრამა, სქემა.

დარგების მიხედვით ერთმანეთისაგან განარჩევენ ვიზუალიზაციის სხვადასხვა სახეს: სამედიცინო ვიზუალიზაცია, ტექნიკური ვიზუალიზაცია, არქიტექტურული ვიზუალიზაცია. ასეთ შემთხვევაში, როგორც წესი, იგულისხმება საგნის 3D-მოცემუ-

ლობის ფორმატში წარმოდგენა. მათგან განსხვავებით სამეცნიერო ვიზუალიზაცია ასახავს არა საგნობრივ, არამედ აბსტრაქტულ მიმართებებს და უნდა აკმაყოფილებდეს სამ აუცილებელ კრიტერიუმს: იგი უნდა იყოს გამომსახველობითი, უნდა ახდენდეს შთაბეჭდილებას და ეყრდნობოდეს რეალურ მონაცემებს, ანუ ვიზუალიზაცია უნდა იყოს **ექსპრესიული, ეფექტური და ადეკვატური.**

ინფორმაციის ვიზუალიზაცია (იგივე მონაცემთა ვიზუალიზაცია) კვლევის ის სფეროა, რომელიც უზრუნველყოფს დიდ მონაცემთა (***big data***) კომპიუტერული საშუალებებით გადამუშავებასა და შედეგების გრაფიკულად წარმოდგენას.

მაგ., იმის დასადგენად, თუ რომელი სიტყვა იხმარება ყველაზე ხშირად მოცემულ ტექსტში (მაგ., „შუშანიკის წამებაში"), შეგვიძლია გამოვითვალოთ ტექსტში სიტყვაფორმების სიხშირის კოეფიციენტი და შედეგები წარმოვადგინოთ ტაბულის სახით. ვიზუალიზაციის პროგრამული საშუალებები კი საშუალებას გვაძლევს კვლევის შედეგები ვიზუალიზაციის გამოყენებით გავათვალსაჩინოოთ.

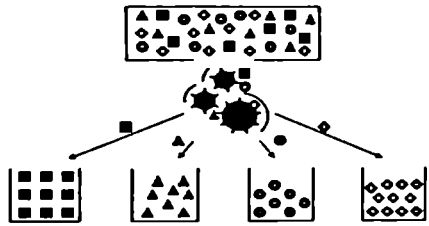
ვიზუალიზაციის დროს გამოიყენება ისეთი დარგებისა და დისციპლინების მეთოდური ინსტრუმენტები, როგორიცაა ინფორმატიკა, კოგნიტური მეცნიერება, სტატისტიკა, data mining.

Data mining[121], როგორც მეთოდი, წარმატებით გამოიყენება მონაცემთა დიდი ბაზების მანქანური დამუშავების დროს. ცნებაში **data mining** იგულისხმება ბაზაში მსგავს ობიექტთა ჯგუფების (კლასტერების) გამოხშირვა-დახარისხება ახალი ნიმუშების გამოვლენის მიზნით. ამგვარად შერჩეული ერთგვაროვანი კლასტერები ერთიანდებიან იერარქიულ (დაქვემდებარებულ) ან აგლომერაციულ (თანაბარდონიან) სისტემაში. აღნიშნული პროცესი მიზნად ისახავს ცოდნის მოპოვებას იმის შესახებ, თუ რა არის სტატისტიკური თვალსაზრისით დაჯგუფებადი, რომელი ჯგუფია ჯერ კიდევ უცნობი და ამდენად, აღუწერელი და რომელია გამოსადეგი იმისათვის, რომ დავადგინოთ რეგულარული მიმართებები, ურთიერთგანმსაზღვრავი წესები და ჯერ კიდევ

---

[121]. ამ ცნების ქართული თარგმანი ჯერჯერობით არ გაგვაჩნია, თუმცა ეს მარტო ქართულის პრობლემა არ არის. მაგ.: Data mining-ის გერმანულ შესატყვისად გამოიყენება მისი გერმანული დაწურილობა Data-Mining. აღნიშნული ცნების შემოთავაზებული ვარიანტები გერმანული ენისათვის, როგორიცაა *"Datenmustererkennung" და "Datenschürfung"* სპეციალისტების მიერ უარყოფილ იქნა, ვინაიდან ვერ გადმოსცემს მისი, როგორც ფენომენის უმთავრეს თვისებას (ამოცანის არა სისტემატიზებული, არამედ ახალი ნიმუშები), ან სიტყვასიტყვითი თარგმანია, რომელშიც დაიკარგა პროცესის შინაარსობრივი მხარე. არც ლუღენის ლექსიკონში შესული ტერმინი *"Datenförderung"* არ არის გაზიარებული სპეციალისტების მიერ.

დაუდგენელი ურთიერთდამოკიდებულებები საანალიზო ელემენტების გროვაში. ანალიზის ამ სახეს **კლასტერული ანალიზი (cluster analysis / clustering)** ეწოდება.



კლასტერული ანალიზის ვიზუალიზაცია

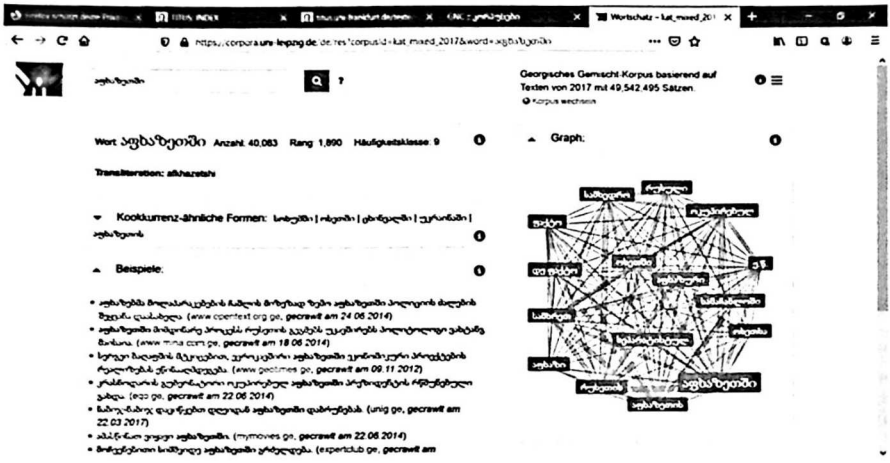
Data mining-ის მიზანია ახალი ჯგუფების გამოვლენა და არა წინასწარ მოცემული ჯგუფების კლასიფიკაცია. მისი გამოყენება ჰიპოთეზების გენერირების საშუალებას იძლევა, რომელიც შემდეგ ვერიფიკაციის შედეგად იხვეწება და ზუსტდება.

Data mining–ის პროცესის საწყისი და ერთ-ერთი უმნიშვნელოვანესი ეტაპია მონაცემთა ბაზის გადამოწმება **ორგანული ან სტრუქტურული შეუსაბამობის (ინკონსისტენტურობის)** თვალსაზრისით. საწყის ეტაპზე, როგორც წესი, მიმდინარეობს ობიექტთა დაჯგუფება და არატიპიური შემთხვევების გამოვლენა. მომდევნო ნაბიჯებია: **კლასტერული ანალიზი, კლასიფიკაცია** და **განპირობებულობის დადგენა**. ეს უკანასკნელი მიზნად ისახავს **კორელაციის წესების** დადგენას, ანუ რა პირობებში არის მოსალოდნელი ორი ან მეტი ობიექტის თანამყოფობა.

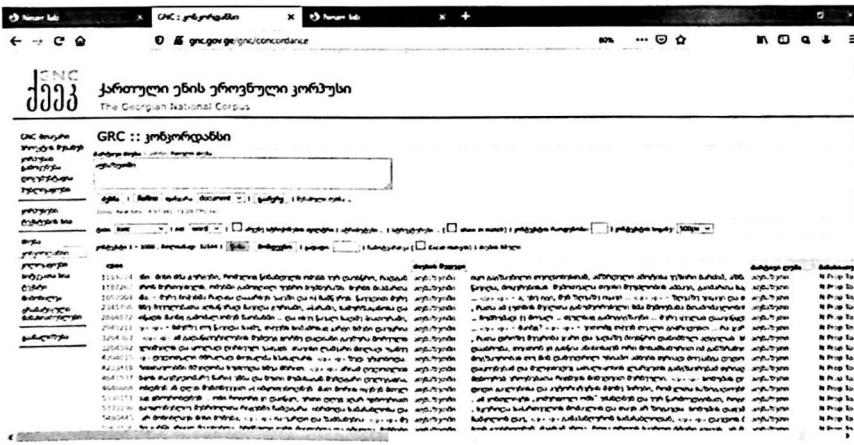
კორპუსლინგვისტიკაში ვიზუალიზაციის გამოყენების ერთ-ერთ ეფექტურ შედეგს წარმოადგენს ლექსიკურ ერთეულთა ტოპოლოგიური და ფუნქციური თანაარსებობის – **კოლოკაციისა** და **კოოკურენციის** – კვლევა. აღნიშნულ ტერმინებთან დაკავშირებით, ქართულ სამეცნიერო სივრცეში არსებულ დისკუსიიდან გამომდინარე, საგანგებოდ გვინდა აღვნიშნოთ, რომ მართალია შესაძლებლად მიგვაჩნია აღნიშნული ტერმინების „გაქართულება“, მაგალითად **თანამგზავრი** *vs.* **თანმხლები სიტყვა**, მაგრამ ვინაიდან ტერმინი **კოლოკაცია/კოლოკანტი** ქართულ სამეცნიერო ენაში დამკვიდრებული ტერმინია, არასწორად მიგვაჩნია მისი მუწყვილის – **კოოკურენცია/კოოკურენტის** „გაქართულება“.

ლაიფციგის უნივერსიტეტის მიერ შექმნილ მრავალენოვანი კორპუსი საშუალებას გვაძლევს არა მარტო მოვიძებნოთ ნებისმიერი სიტყვა კორპუსში, არამედ მოვიძიოთ მოცემული სიტყვაფორმის კონტექსტი კოლოკანტებისა და კოოკურენტების

გათვალისწინებით. ქვემოთ მოცემულია სიტყვაფორმის „აფხაზეთში“ კოოკურენტების ვიზუალიზაცია Wordschatz-ის ქართული ენის კორპუსის მაგალითზე[122].



ვიზუალიზაციის მიხედვით ცალსახად ჩანს, თუ რა ენობრივ გარემოში გამოიყენება ეს სიტყვა, რაც ამ რეგიონში ძალადობაზე და კონფლიქტზე მიუთითებს. ეს სიტყვაფორმებია: *სეპარატისტული, რუსეთის, აფხაზეთის, სამხრეთსა, ოსეთის, აფხაზი, სამაჩაბლოში, აფხაზური, დე ფაქტო, ე.წ. ოსეთში, ოკუპირებული, სამხედრო, რუსული.*



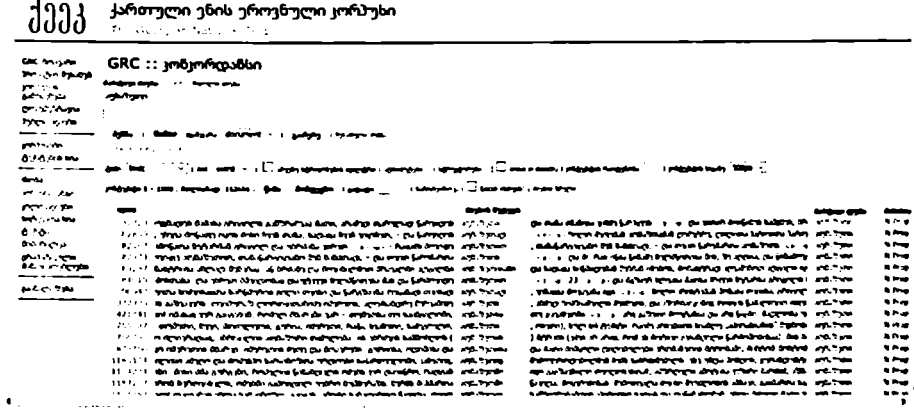
როგორც ვხედავთ, ქართული ენის ეროვნულ კორპუსში სიტყვაფორმა „აფხაზეთში“ 32544-ჯერ იძებნება. ქვემოთ წარმოდგენილი კონკორდანსი მარტივი ძიების

122 https://corpora.uni-leipzig.de/de?corpusId=kat_mixed_2017&word=#

შედეგია: საძიებო სისტემა კორპუსში მხოლოდ აღნიშნულ სიტყვაფორმას (ტოპონიმს *აფხაზეთი*) -ში თანდებულთან კომბინაციაში ეძებს.

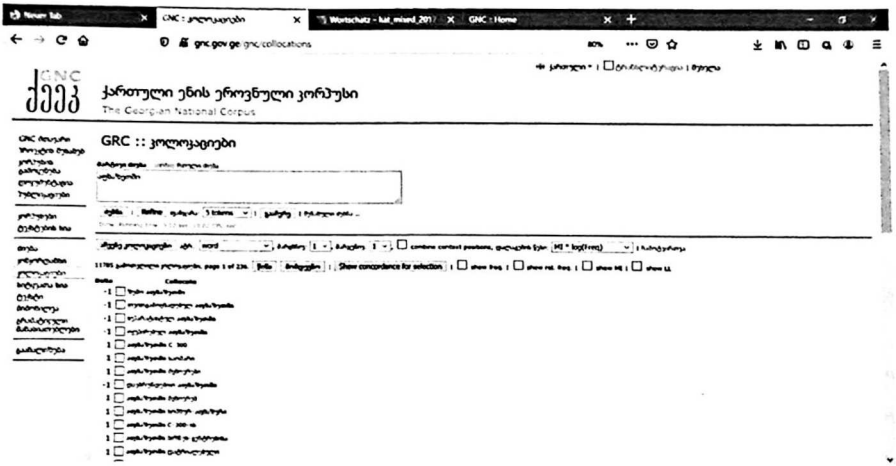
თუმცა, ქართული ენის კორპუსი საშუალებას იძლევა, აღნიშნული ტოპონიმის ყველა შესაძლო ფლექსიური ფორმა მოძებნოს – კორპუსმენეჯერი გრამატიკული წესების ბანკის მონაცემებზე დაყრდნობით ახორციელებს „აფხაზეთ“ ფუძის შემცველი სიტყვაფორმების ავტომატურ ანალიზს და კორპუსში პოულობს მას როგორც სხვადასხვა ბრუნვასა და რიცხვში, ასევე თანდებულიან და ნაწილაკიან ფორმებსაც.

ძიების ეს შესაძლებლობა საგანგებო ოპერაციით – **რთული ძიების** საშუალებით ხორციელდება. როგორც ქვემოთ მოყვანილ სურათზე ჩანს, მიუხედავად იმისა, რომ საძიებო ველში სიტყვაფორმა „აფხაზეთი“ არის მიცემული, კორპუსის მენეჯერი კორპუსში ეძებს ყველა შესაძლო ფორმას: *აფხაზეთს, აფხაზეთად, აფხაზეთით, აფხაზეთისარი, აფხაზეთის, აფხაზეთისა, აფხაზეთიც, აფხაზეთში* და ა.შ. რთული ძიების პროცესი ეფუძნება ლემის განსაზღვრის პრინციპს: კორპუსმენეჯერი კორპუსში ეძებს ლემის „აფხაზეთი“ რეალიზების ყველა შესაძლო ვარიანტს – სხვადასხვა ბრუნვის ფორმებს თანდებულიანი და ნაწილაკიანი ფორმების ჩათვლით.

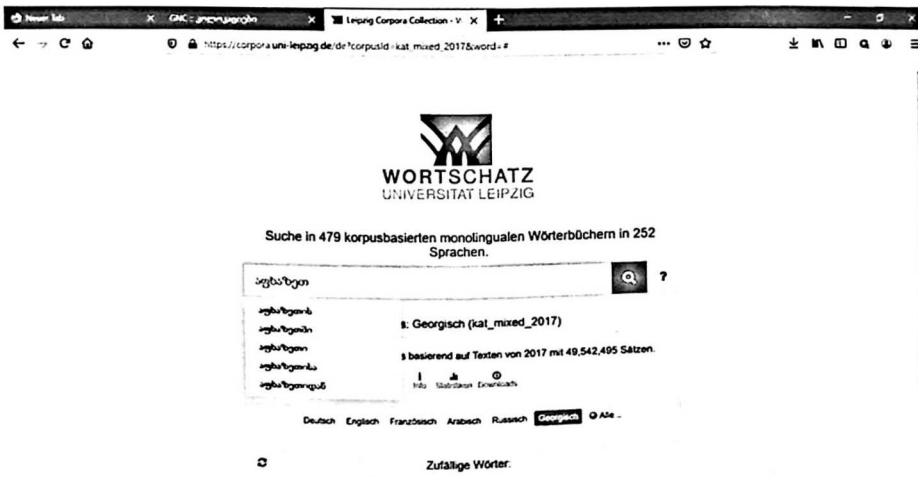


ძიების შედეგად გენერირებული კონკორდანსის გარდა ქეგკ-ში ასევე შესაძლებელია **კოლოკაციების** გენერირება. კოლოკაციების თვალის გადაელებაც კი ნათლად წარმოაჩენს საქართველოს ამ რეგიონში არსებული კონფლიქტის ენობრივ რეალიზაციას ლექსიკის თვალსაზრისით: აქ გამოყენებული კოლოკაციებია: თვით-

გამოცხადებული, სეპარატისტულ, ოკუპირებულ, საომარი, მცხოვრები, დავბრუნდებით და ა.შ.



რთული ძიების განხორციელება არ არის გათვალისწინებული ლაიფციგის უნივერსიტეტის ქართული კორპუსის საძიებო სისტემაში. როგორც ქვემოთ მოყვანილი სურათიდან ხედავთ, საძიებო სისტემა საძიებო სიტყვის კორპუსში არსებული ვარიანტების ჩამონათვალს გვთავაზობს, რომელთაგან მხოლოდ ერთი შეგვიძლია ავირჩიოთ:



თუმცა, აქვე გვინდა აღვნიშნოთ, რომ კოოკურენტების ძიებისა და შედეგების ვიზუალიზაციის თვალსაზრისით ლაიფციგის ქართული კორპუსი სხვა კორპუსებთან შედარებით გაცილებით შთამბეჭდავია.

8.6. კოლოკაცია და კოოკურენცია

ცნება **კოლოკაცია (collocation)** ტექსტის ანალიზის ერთ-ერთი ძირითადი ცნებაა და ნიშნავს ორი ლექსემის უშუალო თანაარსებობას, გვერდიგვერდ პოვნიერებას. ენათმეცნიერებაში **კოლოკაცია**, როგორც ცნება, პირველად **ჯონ რუპერტ ფერთმა** (John Rupert Firth) გამოიყენა მოცემულ კონტექსტში ლექსემების მნიშვნელობის დასაზუსტებლად, თუმცა აღნიშნული ცნება დღეისათვის მეტნაკლებად დიფერენცირებულად გამოიყენება ენის სხვადასხვა დონეზე:

კოლოკაცია აღნიშნავს ლექსემათა თანაარსებობას, რომელიც განპირობებულია:

ა) არსობრივი ურთიერთგანპირობებულობით დენოტატებს შორის: *„ყვავი→ შავი“*;
ბ) სემანტიკური ასოციაციით: *„დღე ↔ ღამე“*;
გ) ლექსიკური სოლიდარობით: *„ქერა → თმა“*;
დ) ფრაზეოლოგიზმებით: *„განაჩენის გამოტანა“*, *„თავბედის დაწყევლა“*;
ე) კონცეპტუალური სტერეოტიპებით *„ტრაგიკულად დაიღუპა“*, *„ბრწყინვალე კარიერა გაიკეთა“*.

კორპუსლინგვისტიკაში კოლოკაცია აღნიშნავს ყველაზე ხშირად გამოყენებულ, სტატისტიკურად დაანგარიშებად ლექსიკურ ელემენტთა (კორპუსში – ტოკენების) თანაარსებობას. ქართული ენის ეროვნულ კორპუსში, მაგ., სიტყვაფორმა ***„ღამე“*** 9872 – ჯერ დასტურდება. ***„ღამის“*** კოლოკანტები (კრიტერიუმებით **„ინფორმაციულობა“ + “სიხშირე“**) კი ასე გამოიყურება[123]:

1. *მთელი ღამე (1457 შემთხვევა)*
2. *მზიანი ღამე (110 შემთხვევა)*
3. *ბნელი ღამე (66 შემთხვევა)*
4. *მთვარიანი ღამე (33 შემთხვევა)*
5. *წუხანდელი ღამე (30 შემთხვევა)*
6. *უკუნი ღამე (19 შემთხვევა)*

რაც უფრო მაღალია კოლოკაციის სიხშირე, მით უფრო მეტია იმის ალბათობა, რომ საქმე გვაქვს მყარ შესიტყვებასთან. მაგ.: ზედსართავი სახელის ***„სუსტი“***

[123] სტატისტიკური მონაცემების თარიღი: 4.01.2021.

კოლოკანტებიდან (სულ კოლოკაციის 3365 მაგალითი) ყველაზე ხშირად ხმარებულია *სუსტი წერტილი* (252), *სუსტი სქესის* [წარმომადგენელი] (190), *სუსტი რგოლი* (82).

კოლოკაცია ენობრივი სპეციფიკის ამსახველია და მისი პოვნიერება ენაში განპირობებულია ისეთი ფაქტორებით, როგორიცაა სემანტიკური თავსებადობა, ლოგიკური განპირობებულობა, ფრაზეოლოგიზმები და სტერეოტიპები. არსებობს უნივერსალური კოლოკაციები, თუმცა ენები, როგორც წესი, ამ თვალსაზრისითაც განსხვავდებიან ერთმანეთისაგან: შდრ. *ნიშნის დაწერა* (ქართ. "*მასწავლებელმა ცუდი ნიშანი დაუწერა*") და *ნიშნის მიცემა* (გერმ. Der Lehrer gab ihn eine schlechte Note შდრ. "*მასწავლებელმა ცუდი ნიშანი მისცა*").

ხშირად სემანტიკურად ახლოს მდგომი ლექსემები საგრძნობლად განსხვავდებიან კოლოკაციების თვალსაზრისით: შდრ., მაგ., *მიზნის მიღწევა*, მაგრამ *შედეგის მიღება*. კოლოკაციის ფენომენს ვალტერ პორციგი (Walter Porzig) ისეთი ცნებების საშუალებით აღწერდა, როგორებიცაა **არსობრივი შინაარსობრივი კავშირი** და მნიშვნელობის (საერთო) სინტაქსური არე, ეუგენი კოსარიუსთან (Eugenio Coseriu) კი მას **ლექსიკური სოლიდარობა** ეწოდება.

კოოკურენტი (co-occurring term) კოლოკაციის საგანგებო შემთხვევაა და აღნიშნავს ლექსემათა ლოგიკურ-კოჰერენტულ რელაციას.

კორპუსლინგვისტიკაში ცნება **კოოკურენცია** აღნიშნავს ორი ტოკენის ისეთ თანაარსებობას კონტექსტში, რომელიც სიხშირის მაღალ კოეფიციენტს გვიჩვენებს. კოოკურენცია ეყრდნობა ე. წ. Zipf-ის კანონს: თუ $f * r \approx k$ (f ~ ერთი ტოკენის სიხშირე კორპუსში, r ~ რანგი ნუმერირებულ სიაში), მაშინ $f \approx k / r$, ანუ კორპუსში სიტყვის სიხშირე უკუპროპორციულია მის რანგთან სიხშირის სიაში.

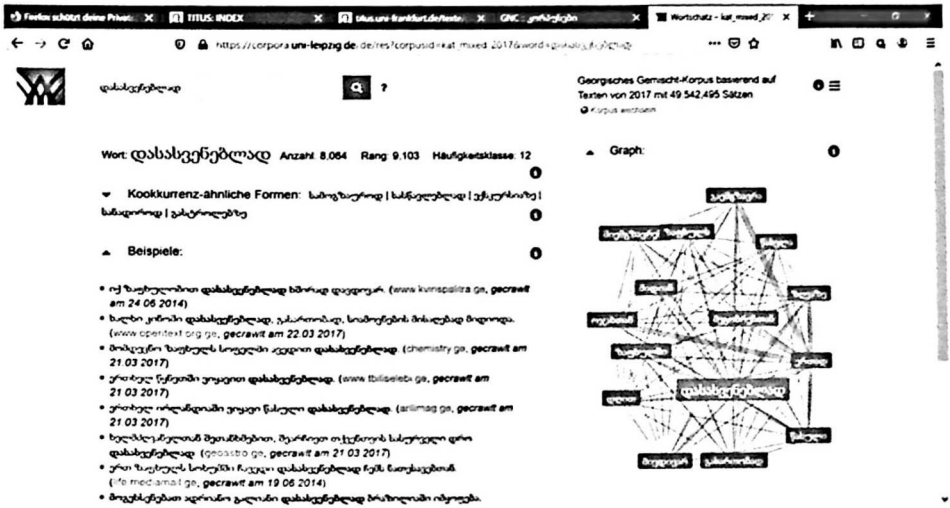
8.7. ინფორმაციის ვიზუალიზაცია გერჰარდ ჰაიერის Wortschatz-ში

ინფორმაციის ვიზუალიზაციის თეორია („განწყობების წაკითხვა") წარმატებით გამოიყენა ლაიფციგის უნივერსიტეტის პროფესორმა გერჰარდ ჰაიერმა, რომელმაც საგანგებო მულტილინგვური პლატფორმა შექმნა ინფორმაციის ვიზუალიზაციის მიზნით – ენობრივ მონაცემთა ვიზუალიზაცია **Wortschatz** -ში 230 ენის[124] პრესის მასალებს მიემართება.

---

[124] მათ შორის ქართული პრესის მასალებიც: ქართული პრესის ელექტრონული არქივი Opentext (http://www.opentext.org.ge) და ახალი ამბების სააგენტო Civil Georgia სამენოვანი პორტალით (http://www.civil.ge).

გ. ჰაიერის პროგრამა ეფუძნება კოლოკაციისა (ტექსტში სიტყვის მეზობლად მდგარი მეორე სიტყვა ერთი ან ორივე მიმართულებით) და კოოკურენციის (ტექსტურ მასივში მოცემულ სიტყვასთან ყველაზე ხშირად გამოყენებული მეორე სიტყვა) ცნებას.



როგორც ვიზუალიზაციის ზემომოყვანილი მაგალითიდან ჩანს, წინადადებაში *„ზაფხულში დასასვენებლად ზღვაზე მივდივარ“* სიტყვის *„დასასვენებლად“* კოლოკანტებია *„ზაფხულში“* და *„ზღვაზე“*. მაშინ როდესაც, ამ სიტყვის კოოკურენტებად (ყველაზე ხშირად ხმარებული სიტყვებად) **Wortschatz** – ის მიხედვით გვევლინება *„ვიყავი“* და *„ოჯახთან ერთად“*.

Wortschatz–ის პლატფორმა, როგორც ზემოთ უკვე აღვნიშნეთ, 230 ენის რესურსებს მიემართება. ამგვარი მრავალენოვანი კორპუსების არსებობა საშუალებას გვაძლევს საგანგებოდ ვიკვლიოთ ლექსემათა კოოკურენტების ტიპოლოგია სხვადასხვა ენაში უნივერსალიების დადგენის მიზნით. მაგ.: ქართულის შემთხვევაში ტექნიკური ცნების „დემონტაჟი“ კოოკურენტებია სიტყვები: *რეჟიმის* (111.33), *მემორიალის* (58.79), *ძეგლის* (56.04), *ხელისუფლების* (48.56), *სახლის* (43.91), *შენობის* (33.47). საგულისხმოა, რომ Wortschatz–ის პლატფორმის მიხედვით, მრავალკომპონენტიან კოოკურენტებში ლიდერობს „*დემონტაჟის*“ მარცხენა კოოკურენტი „*საბავშვოს ხელისუფლების*“. Wortschatz-ის გერმანული რესურსების მიხედვით შესაბამისი ლექსემის

Demontage *კოოკურენტებია: ზეჰოფერი*[125], *ძეგლი* (Denkmal), *ვესტერველე*[126]), *სოციალისტური სახელმწიფოს* (Sozialstaates), *სოციალისტური სისტემის* (Sozialsystems). Wortschatz–ის პლატფორმა ძირითადად პრესის მასალებს მიემართება, ამიტომ გასაკვირი არაა, რომ საკვლევი ლექსემის *„დემონტაჟი“* კოოკურენტები, სემანტიკურად თავსებადი პარტნიორი სიტყვების – *შენობა, ძეგლი, კონსტრუქცია* – გარდა, პოლიტიკური პირები ან პოლიტიკური კონტექსტიც გახლავთ.

ინფორმაციის ვიზუალიზაცია წარმატებით გამოიყენება კვლევებისათვის ლიტერატურათმცოდნეობაში, სოციოლოგიაში, პოლიტოლოგიაში და ა.შ. მაგ., ეფექტიანად შეიძლება ვიზუალიზაციის გამოყენება არა მხოლოდ ცალკეული სიტყვების, არამედ მთელი ციტატების შემთხვევაშიც. უდავოდ საინტერესო უნდა იყოს ქრონოლოგიურ ჭრილში ეპოქების მიხედვით ისეთი საკითხის კვლევა, როგორიცაა, მაგალითად, **ეთნიკური იდენტიფიკაციის კომპონენტური ანალიზი** ანუ ეთნიკური იდენტიფიკაციის პოლიტიკური, რელიგიური, გეოგრაფიული, ეთნოლოგიური, ლინგვისტური საფუძვლები: როდის აქტიურდებოდა ქართველთა ცნობიერებაში ისეთი ციტატებით ოპერირება, როგორიცაა *„ქართლად ფრიადი ქუეყანაჲ აღირაცხებისჲ“* (გრიგოლ ხანძთელი) ან *„ვითარმედ ყოველი საიდუმლოჲ ამას ენასა შინა დამარხულ არს“* (იოანე ზოსიმე) და სხვა. კორპუსში მონაცემთა ვიზუალიზაციის ამგვარი შესაძლებლობები უდავოდ დამაჯერებელ ემპირიულ საფუძველს უქმნის სოციოლოგებს, პოლიტოლოგებს, ისტორიკოსებსა და სხვა დარგის სპეციალისტებს.

ინფორმაციის ვიზუალიზაცია (საზოგადოებრივი აზრის მაჯისცემის მოსმენა) არა მხოლოდ კვლევის, არამედ კონფლიქტების პროგნოზირების ეფექტიანი საშუალებაც გახლავთ. მეტიც, არსებობს კოოკურენტების „მართვის“ ონტოლოგია. ვინაიდან კოოკურენცია ძირითადად გარკვეული სიტყვების ერთდროული ხმარების სიხშირეზეა დამოკიდებული, იგი ადვილად მანიპულირებადია. მიზანმიმართული მცდელობით შესაძლებელია მისი სიხშირის გაზრდა (სასურველი კოოკურენტების ხშირი გამოყენება) და შესაბამისად, ამა თუ იმ მოვლენისადმი პერსპექტივის შეცვლა. ამის საუკეთესო მაგალითი გახლავთ ნიუ-იორკ ტაიმსის ელექტრონული კორპუსიდან გამოთვლილი კოოკურენციების სურათი ცნებით „აბუ-გრეიბი[127]“. გამოკვლევის პირველ ქრონოლოგიურ ფაზაში ამ სიტყვის ირგვლივ დაჯგუფდა ისეთი ცნებები, რო-

---

[125] ჰორსტ ზეჰოფერი (Horst Seehofer), გერმანელი პოლიტიკოსი, 2008 წლიდან პოლიტიკური პარტიის CSU თავმჯდომარე, ბავარიის მინისტრთა საბჭოს თავმჯდომარე.

[126] გვიდო ვესტერველე (Guido Werterwelle), გერმანელი პოლიტიკოსი, 2009-11 წლებში გერმანიის ვიცეკანცლერი, 2009-13 წლებში გერმანიის საგარეო საქმეთა მინისტრი.

[127] Abu-Ghraib - ერაყში არსებული ციხე ამერიკული ჯარების მმართველობის ქვეშ.

გორიცაა *„წამება“, „ადამიანის უფლებათა დარღვევა“, „არაჰუმანურობა“* და ასე შემდეგ. მეორე ქრონოლოგიურ ფაზაში *„აბუ-გრეიბ“*-თან ასოცირებადი კოოკურენციები მოიცავდა ცნებებს *„ამერიკის უსაფრთხოება“, „ტერორისტი“, „თერთმეტი სექტემბერი“* და ასე შემდეგ. ალბათ შეიძლება ვივარაუდოთ, რომ კოოკურენციის თემატური დისლოკაცია ამ სიტყვასთან (resp. ფაქტთან) მიმართებაში ამერიკის ხელისუფლების მიზანმიმართული მონდომების შედეგი იყო.

9. მეთოდი კორპუსლინგვისტიკაში

კორპუსლინგვისტიკის ერთ-ერთ ძირითად ამოცანას წარმოადგენს კვლევის საკუთარი ინსტრუმენტების განვითარება – **კორპუსის აგების, მართვისა და გამოყენების მეთოდოლოგიის** შექმნა, ერთი მხრივ, და **კორპუსლინგვისტური კვლევის მეთოდური ინვენტარის** შემუშავება, მეორე მხრივ.

კორპუსლინგვისტიკა არა მხოლოდ საკუთარი მეთოდებით აღჭურვილი დარგია, არამედ კვლევის საკუთარი ინსტრუმენტებიც გააჩნია. მართალია, იგი ხშირად იყენებს ტრადიციული ლინგვისტიკისა და ინფორმატიკის, როგორც დამოუკიდებელი დარგების მეთოდებს, მაგრამ ისევე, როგორც სხვა ინტერდისციპლინური დარგები, მასაც გააჩნია საკუთარი მეთოდური ინვენტარი, რომელიც სრულიად ახალ, საგანგებოდ კორპუსლინგვისტიკისათვის შექმნილ მეთოდურ რესურსს წარმოადგენს.

ტრადიციული ენათმეცნიერების წარმომადგენლები კორპუსლინგვისტიკას არა დამოუკიდებელ დისციპლინად, არამედ ლინგვისტიკის ერთ-ერთ მეთოდად მიიჩნევენ, რომელიც საშუალებას აძლევს ლინგვისტს კვლევა თანამედროვე ტექნოლოგიური საშუალებებით აწარმოოს. „თანამედროვე ტექნოლოგიურ საშუალებებში“ კომპიუტერის გამოყენება იგულისხმება, თავად კომპიუტერი კი საჭირო ენობრივი მონაცემის „სწრაფ მოძებნასთან“ ან „სწრაფ წედომასთან“ ასოცირდება. ამ კონტექსტში სრულიად უგულებელყოფილია მონაცემებთან მიმართებაში ტექნიკური რევოლუციის ერთი უპირატესობა: **ციფრულ მონაცემთა სისტემური კვლევა.**

მონაცემთა ახალი, ციფრული ფორმატის შექმნას ორი არსებითი მნიშვნელობა ჰქონდა ენათმეცნიერებისათვის:

1) მონაცემთა „გაციფრებამ“ შესაძლებელი გახადა დროსა და სივრცეში დაშორებული მონაცემების ერთიან, ვირტუალურ სივრცეში განთავსება;

2) შეიქმნა საგანგებო ტექნოლოგიები ენობრივი რესურსების საგანგებო მათემატიკურ-ლოგიკური წესებისა და ალგორითმების საშუალებებით გადამუშავების განსახორციელებლად.

აქ არსებითად მიგვაჩნია ლინგვისტიკის, როგორც მეცნიერების კონკრეტული დარგის, განვითარების დიალექტიკის გათვალისწინება, ანუ იმის აღნიშვნა, რომ **ემპირიული მასალის სისტემატიზაციაზე ორიენტირებულმა ტრადიციულმა ლინგვისტიკამ** მე-20 საუკუნეში თვისებრივად ახალი, თეორიაზე ორიენტირებული **გენერატიული ლინგვისტიკა** წარმოშვა, რომელიც კვლევის პროცესში ძირითადად იყენებდა არა ინდუქციის (ემპირიული ფაქტების განზოგადების მეთოდი), არამედ დედუქციის მეთოდს (ლინგვისტური პოსტულატების გენერირების და მათი ემპირიულ მასალაზე გადამოწმების მეთოდი). მონაცემთა ახალი ტიპის, ციფრული მონაცემების გაჩენამ ერთმანეთს კვლავ შეახვედრა თეორია და ემპირია, ოღონდ არა როგორც ორი დამოუკიდებელი კომპონენტი, არამედ როგორც ერთიანი, ჰომოგენური მთლიანობა.

ლინგვისტიკის განვითარების დიალექტიკა, აქედან გამომდინარე, ჩვენ ასე წარმოგვიდგენია: **პრაქტიკული ემპირიიდან** (ემპირიული მონაცემების ანალიზი) ლინგვისტიკა იქცა **„ჰიპოთეზურ“ თეორიად** (ჰიპოთეზურ პოსტულატებზე აგებულ თეორიად), რომელმაც **ემპირიული თეორია** (ემპირიულ მასალაზე ვერიფიცირებადი თეორია) შვა.

ერთ-ერთ უმნიშვნელოვანეს განსხვავებას ამ სამ საფეხურს შორის ქმნის ემპირიული რესურსების მოცულობა – ტრადიციული ლინგვისტიკისაგან განსხვავებით თანამედროვე ენათმეცნიერება მონაცემთა დიდი მოცულობის ბაზებს (Big Date) ეყრდნობა. **დიდი ციფრების კანონმა** სრულიად შეცვალა ენობრივი სისტემის მოზაიკაში ცალკეული ემპირიული შემთხვევების ადგილი და მნიშვნელობა. ენობრივი შემთხვევა, რომელიც თუნდაც ერთხელ დასტურდება 130 მილიონი სიტყვაფორმის მომცველ კორპუსში, ისეთივე მნიშვნელოვანია, როგორც ნებისმიერი მაღალი სიხშირით აღჭურვილი ენობრივი ფაქტი.

პრობლემის მეთოდოლოგიური ასპექტი ერთ-ერთ უმნიშვნელოვანეს საკითხს წარმოადგენს კორპუსლინგვისტიკის, როგორც დამოუკიდებელი ინტერდისციპლინური დარგის ლეგალიზაციისათვის და მისი არსებობის შესახებ დისკუსიის ეფექტურობაც სწორედაც რომ ამ საკითხის კორექტულად გადაწყვეტაზეა დამოკიდებული.

9.1. კორპუსზე დამყარებული კვლევა

ტრადიციული ლინგვისტიკა ემპირიული ფაქტების ანალიზს ემყარებოდა. კორპუსლინგვისტიკაც ემპირიულ ფაქტებს ეყრდნობა, ოღონდ გაცილებით უფრო დიდი მოცულობის ფაქტებს. გადაჭარბებული არ იქნება თუ ვიტყვით, რომ თანამედროვე თეორიული ლინგვისტიკა არსებითად კორპუსულ კვლევებს იყენებს – დღეს

ცნება „კვლევა“ ცალსახად „კორპუსზე ორიენტირებულ კვლევას“ გულისხმობს, ხოლო გრამატიკა – „კორპუსულ გრამატიკას“[128].

თანამედროვე ენათმეცნიერება კორპუსთან მიმართებაში ორი სახის კვლევას იყენებს: კორპუსზე დამყარებულ კვლევებს და კორპუსით განპირობებულ კვლევებს.

კორპუსზე დამყარებულ კვლევაში **(Corpus-based studies)** იგულისხმება მეთოდი, რომელიც საშუალებას იძლევა შევამოწმოთ ლინგვისტური ჰიპოთეზის ვალიდურობა კორპუსზე დაყრდნობით. ხაზგასასმელია ის ფაქტი, რომ ჰიპოთეზის ჩამოყალიბება ამ შემთხვევაში არ უკავშირდება კორპუსულ კვლევას; კორპუსის ბაზაზე ხდება მისი სისწორისა და დეტალურობის ხარისხის შემოწმება. კორპუსზე დამყარებული კვლევის შედეგად ვახდენთ წესის ან ჰიპოთეზის დადასტურებას (ვერიფიცირება) ან უარყოფას (ფალსიფიცირება). თვალსაჩინოებისათვის განვიხილოთ ერთი მაგალითი – „ლიმონი“. ქართული ენის განმარტებით ლექსიკონში *ლიმონი* დახასიათებულია, როგორც ფუძეკუმშვადი სახელი – ფრჩხილებში მითითებულია ნათესაობითში შეკუმშული ფორმა: *ლიმონ-ი (ლიმნისა)*. თუ საილუსტრაციოდ მოტანილ მაგალითებს გადავაველებთ თვალს, აღმოვაჩენთ, რომ მაგალითებში შეკუმშული ფორმის – *ლიმნის* – გვერდით ერთხელ მხოლობითი რიცხვის შეუკუმშავი ფორმაა დადასტურებული (*ლიმონის ხე*), ერთხელ კი მრავლობითი რიცხვის შეუკუმშავი ფორმა (*ფცქვნიდა ლიმონებს...*).

ცხადია, რომ განმარტებითი ლექსიკონის მიხედვით ცალსახად ვერ ვუპასუხებთ კითხვას, კუმშვადია თუ არა „*ლიმონი*“ ქართულში.

ქართული ენის ეროვნულ კორპუსში ამ ლექსემასთან დაკავშირებით შემდეგი სურათი გვაქვს: *ლიმონი (5), ლიმონმა (3), ლიმონს (50), ლიმნის (38) შდრ. ლიმონის (157), ლიმნით (3) შდრ. ლიმონით (27), ლიმონად (5), ლიმნები (3) შდრ. ლიმონები (23), ლიმონებს (6), ლიმნების (1) შდრ. ლიმონების (4), ლიმნებით (1) შდრ. ლიმონებით (1).*

კორპუსში დადასტურებული ფორმების სტატისტიკური ანალიზი ცალსახად გვიჩვენებს ტენდენციას ქართულში სიტყვა *ლიმონი* უკუმშველი ფუძის ფორმით

---

[128] ნაშრომის დასტურად გამოდგება, მაგალითად, ფრანტისეკ შტიხას (Frantisek Štícha) მიერ დაარსებული საერთაშორისო კონფერენციების სერია „გრამატიკა და კორპუსები“. საერთაშორისო კონფერენციის ეს სერია 2005 წელს ჩატარდა პირველად ჩეხეთში და 2009 წლიდან გადმოინაცვლა გერმანიაში. თავდაპირველად სლავურ ენებზე ფოკუსირებული კონფერენცია სულ მალე იქცა კორპუსული გრამატიკის სადისკუსიო ფორუმად, სადაც სხვადასხვა ენის კორპუსული კვლევების რელევანტური შედეგები არის სამეცნიერო დისკუსიის ობიექტი.

გამოიყენოს. კორპუსული კვლევების მიხედვით არაერთარი საფუძველი არ გაგვაჩნია *ლიმონი* ცალსახად კუმშვად არსებით სახელად მივიჩნიოთ, როგორც ეს ქართული ენის განმარტებით ლექსიკონშია მითითებული.

9.2. კორპუსით განპირობებული კვლევა

კორპუსით განპირობებული ლინგვისტური კვლევა (**Corpus-driven studies**) გულისხმობს მეთოდს, რომლის დროსაც ლინგვისტური ჰიპოთეზის ჩამოყალიბება იმთავითვე უკავშირდება კორპუსს, ანუ მეცნიერული პრობლემის დასმის წყაროს წარმოადგენს კორპუსში წარმოებული კვლევის შედეგების ანალიზი.

კვლევის ეს ორი მეთოდი – **კორპუსზე დამყარებული ლინგვისტური კვლევა** და **კორპუსით განპირობებული ლინგვისტური კვლევა** – ერთმანეთისაგან იმით განსხვავდება, რომ პირველ შემთხვევაში საქმე გვაქვს ჰიპოთეზის ვალიდურობის შემოწმებასთან, მეორე შემთხვევაში კი ჰიპოთეზის გენერირებასთან[129]. ორივე შემთხვევაში შედეგად ვიღებთ **ჰიპოთეზის დადასტურებას (ვერიფიკაცია)** ან **უარყოფას. (ფალსიფიკაცია)**.

კორპუსში ორი სახის ძიება არის შესაძლებელია: ე.წ. **KWIC**–ძიება **(key word in context)** ანუ ცალკეული სიტყვების ძიება კონტექსტში და **KRIC** –ძიება **(key relation in context)** ანუ მიმართებების ძიება კონტექსტში.

KWIC–ძიების შემთხვევაში წარმოიქმნება **კონკორდანსი**. კონკორდანსის ცნება, რომელიც თავდაპირველად ბიბლიათმცოდნეობაში, შემდგომ კი ლიტერატურათმცოდნეობაში გამოიყენებოდა, კორპუსლინგვისტიკის ერთ-ერთ ძირეულ ცნებას წარმოადგენს. **კონკორდანსი** არის ანბანურ რიგზე დალაგებული სიტყვების ან ფრაზების სია და მისი გენერირება კორპუსში წარმოებს სპეციალური ალგორითმული სისტემის, **კონკორდანსერის** მეშვეობით. კონკორდანსის სინონიმებად გამოიყენება აგრეთვე **რეგისტრი** ან **ინდექსი**. კორპუსლინგვისტიკაში გამოყენებული ცნება **კონკორდანსი** იმით განსხვავდება სიტყვების უბრალო ჩამონათვალისაგან, რომ იგი ცალკე ამოღებული სიტყვების ჩამონათვალი კი არ არის, არამედ გვიჩვენებს მის ადგილს იმ კონტექსტში, რომელშიც იგი გვხვდება.

---

[129] ზოგიერთი ავტორი (მაგ., Elena Tognini-Bonelli, *Corpus Linguistics at Work*, John Benjamins Publishing, 2001 2001: 84–85) ამ მეთოდზე დაყრდნობით თვლის, რომ ენობრივი კორპუსი თავად „ქმნის" ენის თეორიას.

კორპუსით განპირობებული კილეების საილუსტრაციოდ განვიხილოთ **ზედ** ელემენტის **KWIC**–ძიების შედეგები[130].

ზედ ელემენტი ქართული ენის ეროვნულ კორპუსში 6620-ჯერ დასტურდება[131]. შესაბამისი კონტექსტების პირველადმა ანალიზმა ცხადყო, რომ **ზედ** ელემენტი მიღებულ კონკორდანსში განსხვავებული ფუნქციებით გამოიყენება. განვიხილოთ რამდენიმე მაგალითი:

*(1) ტახტთან ტაბურეტი მოვაჩოჩე, **ზედ** არყის ბოთლი შემოვდგი.*

*(2) ბრალი დასდეს: შენი წაღებულია ხერხი, **ზედ** შენი თითის ანაბეჭდებიაო...*

*(3) გამჭოლი მზერით მიანიშნებდა პატარა მაგიდისაკენ, რომელიც იქვე, გასასვლელ კართან იდგა და **ზედ** ნაკურთხი პური, ღვინო და ზეთი ეწყო.*

*(4) შეევარდნაძე კი სხვანაირად ფიქრობს, მას წასვლა ჯერ არ უნდა, **ზედ** უნდა დააკვდეს იმ სკამს, როგორც ზვიად გამსახურდია დააკვდა.*

*(5) საქართველოს ალაში **ზედ** ცხვირწინ ჩამომაფხრიწეს.*

*(6) მთაწმინდაზე გაუთხარეს სამარე, **ზედ** ილიას გვერდით.*

პირველ ორ მაგალითში **ზედ** დამოკიდებული წინადადების საწყის პოზიციაში დგას და მიემართება მთავარი წინადადების ერთ-ერთ წევრს – პირველ წინადადებაში *ტაბურეტს*, მეორეში – *ხერხს*. ჩვენი მოსაზრების შესამოწმებლად შეგვიძლია გამოვიყენოთ სინტაქსური მოდიფიკაციის წესი (მთავარი და დამოკიდებული წინადადება გარდაქმნათ ორ მარტივ წინადადებად) და ლექსიკური სუბსტიტუციის წესი (**ზედ** ელემენტი ჩავანაცვლოთ მთავარი წინადადების იმ წევრებით, რომელსაც ის მიემართება: ***ტაბურეტი*** და ***ხერხი***):

(1ა) ტახტთან ტაბურეტი მოვაჩოჩე, **ზედ** არყის ბოთლი შემოვდგი >

ა) ტახტთან ტაბურეტი მოვაჩოჩე

ბ) **ტაბურეტზე** არყის ბოთლი შემოვდგი

(2ა) ბრალი დასდეს: შენი წაღებულია ხერხი, **ზედ** შენი თითის ანაბეჭდებიაო>

ა) ბრალი დასდეს

ბ) შენი წაღებულია ხერხი

გ) **ხერხზე** შენი თითის ანაბეჭდებიაო

---

[130] აღნიშნული კილეების შედეგები გამოქვეყნებულია: M. Tandaschwili, *Lokation und Deixis im Georgischen - Funktional-semantische Analyse von zed und šig*. In: PRO GEORGIA, Journal of Kartwelological Studies, N. 23, Warschau;

[131] სტატისტიკური მონაცემის თარიღი: 02.01.2021.

პირველ მაგალითში ორი შესაძლებლობიდან – *ტახტზე* და *ტაბურეტზე* – ლოგიკური შესაბამისობიდან გამომდინარე, მხოლოდ უკანასკნელი ვარიანტია დასაშვები (*ტაბურეტზე*). არჩევანის სიზუსტე ამ შემთხვევაში კონტექსტის ანალიზით არის განპირობებული.

მესამე მაგალითში **ზედ** ელემენტი მიემართება მიმართებით ნაცვალსახელს *„რომელიც“*: რომელიც თავის მხრივ მთავარი წინადადების ერთ-ერთ წევრს – ადგილის გარემოებას (*მაგიდისაკენ*) მიემართება. ჩემოთ გამოყენებული წესების შესაბამისად მესამე წინადადების ტრანსფორმაციის შემთხვევაში მიყიღებთ შემდეგ ვარიანტებს:

(3ა) გამჭოლი მზერით მიანიშნებდა პატარა მაგიდისაკენ, რომელიც იქვე, გასასვლელ კართან იდგა და **ზედ** ნაკურთხი პური, ღვინო და ზეთი ეწყო.

ა) გამჭოლი მზერით მიანიშნებდა პატარა მაგიდისაკენ

ბ) მაგიდა იქვე, გასასვლელ კართან იდგა

გ) *მაგიდაზე* ნაკურთხი პური, ღვინო და ზეთი ეწყო.

აშკარაა, რომ ორივე შემთხვევაში **ზედ** ელემენტი რამდენიმე ფუნქციას ითავსებს, კერძოდ: ა) დეიქტური ფუნქციით გვეელინება, ბ) ანაფორას წარმოადგენს ან გ) დამატებით სუპერესიულობას გამოხატავს – ლექსიკური სუბსტიტუციის შემთხვევაში ჩანაცვლებული სახელი ზე-თანდებულით ფორმდება.

ზედ ელემენტი კატაფორული ფუნქციითაც შეიძლება შეგეხვდეს, როგორც, მაგალითად, მეოთხე წინადადებაში:

4) შეეარდნაძე კი სხვანაირად ფიქრობს, მას წასვლა გვერ არ უნდა, ზედ უნდა დააკვდეს იმ სკამს, როგორც ზვიად გამსახურდია დააკვდა.

ზედ ელემენტი წინ უსწრებს მისამართ სიტყვას და შესაბამისად კატაფორულია.

რაც შეეხება მეხუთე წინადადებებს, აქ **ზედ** ელემენტს სულ სხვა ფუნქცია გააჩნია. **ზედ** ელემენტის პოვნიერების თვალსაზრისით უპირველეს ყოელისა, აღსანიშნავია ორი არსებითი მნიშვნელობის გარემოება:

1) აღნიშნული წინადადება წინა მაგალითებისაგან განსხვავებით მარტივი წინადადებაა;

2) ზედ ელემენტი ელიმინირებადია – მისი ამოღებით წინადადების ფაქტობრივი შინაარსი არ იცვლება.

შეეადაროთ ამ წინადადების ორი ვერსია ერთმანეთს: ერთი **ზედ** ელემენტით, მეორე, **ზედ** ელემენტის გარეშე.

ა) საქართველოს ალაში ზედ ცხვირწინ ჩამომაფხრიწეს.

ბ) საქართველოს ალაში ~~ზედ~~ ცხვირწინ ჩამომაფხრიწეს.

პირველი ვერსია მეორისაგან იმით განსხვავდება, რომ პირველში მთქმელის პირადი დამოკიდებულება, ინტენცია არის მოცემული წინადადებით გადმოცემული ფაქტის მიმართ. *ზედ* ელემენტის ამოღების შემდეგ კი ეს ინტენცია ნეიტრალიზდება – წინადადება "*საქართველოს ალაში ცხვირწინ ჩამომაფხრიწეს*" მომხდარ ფაქტს აღწერს ყოველგვარი მოდალობის გარეშე. *ზედ* ელემენტი, შესაბამისად, მოდალურ ელემენტად უნდა იქნეს მიჩნეული.

ზედ ელემენტის მოდალურობა უკეთესად "იკითხება" მეექვსე წინადადებაში. *ზედ* ელემენტი ამ შემთხვევაში აპოზიციაში დასტურდება (*ზედ ილიას გვერდით)* და გარდა ამისა, აშკარად უარყოფითი კონოტაციის მატარებელია: მთქმელი აღშფოთებულია იმ ფაქტით, რომელიც წინადადებაში არის აღწერილი (მაყანს მთაწმინდაზე გაუთხარეს სამარე, თანაც ილიას გვერდით!). სემანტიკური ფუნქციის თვალსაზრისით, აქ დარღვეულია პრესუპოზიციურობა, რის გამოც *ზედ* ელემენტი მოდალობის, კერძოდ, კონტრაპრესუპოზიციულობის გამომხატავ ელემენტად მიგვაჩნია.

ამგვარად, დასკვნის სახით შეგვიძლია ვთქვათ, რომ *ზედ* ელემენტი ქართულში ფუნქციონირებს როგორც ა) ადგილის ზმნიზედა, ბ) (ფსეუდო)ანაფორა ან კატაფორა, გ) დეიქტური ელემენტი (იწვევს დეიქტურ სპეციფიკაციას) ან დ) მოდალური ელემენტი (გამოხატავს კონტრაპრესუპოზიციულ დადასტურებას).

დასკვნები, რომელიც *ზედ* ელემენტის კორპუსული კვლევის შედეგად მიიღეთ, კორპუსით განპირობებული კვლევის შედეგია. კვლევის ეს შედეგები კიდევ ერთხელ მოწმდება კორპუსში, რათა მოვახდინოთ ჩვენ მიერ დასკვნაში მოცემული ლინგვისტური ჰიპოთეზის ვერიფიცირება.

ზედ ელემენტის სემანტიკური ფუნქცია გარკვეულ კანონზომიერებაშია მის ტოპოლოგიასთან: დეიქტური ფუნქციით, როგორც წესი, ზმნის წინა (ან წინადადების ერთი წევრით დაშორებულ) პოზიციაში გვხვდება, ანაფორის შემთხვევაში დამოკიდებული წინადადების ინიციალურ პოზიციას იკავებს, მოდალური მნიშვნელობით კი სახელის წინ დგას.

დაკვირვების შედეგად აღმოჩენილი კანონზომიერების ვერიფიცირებისათვის გამოვიყენეთ **KRIC**-ძიების მეთოდი შემდეგი რეგულარული გამოსახულებებით:

KRIC:

- **Expression 1.**
    - "ზედ" [ ]? [features="V.*"]

- **Expression 2.**
    - "ზედ" [ ]? [features="N.*"&"!=x"] while x=corresponding suffix
- **Expression 3.**
    - "ზედ" [ ]{0,3} ".*ზე" – 274 contexts

KRIC-ძიების შედეგად გენერირებულმა კონკორდანსმა, ამჯერად საგრძნობლად შემცირებული რაოდენობის ბაზაზე, კიდევ ერთხელ დაგვიდასტურა ჩვენი მოსაზრება **ზედ** ელემენტის მოდალური ფუნქციით გამოყენების შესახებ. საილუსტრაციოდ მოვიტანთ ერთ მაგალითს: *პირამდე ავსებული ურნა ეიღაცას* **ზედ** *ტროტუარზე გადაეპირქვაებინა.*

ზედ ელემენტი, როგორც წესი, **ზე**-თანდებულიანი სახელის წინ დგას, თუმცა რამდენიმე მაგალითი დადასტურდა, სადაც **ზე**-თანდებულის ნაცვლად სხვა თანდებული დასტურდება: *ეახტანგს სახლი* **ზედ** *ტბასთან ედგა.*

აღნიშნული განსხვავება სემანტიკური ფუნქციის თვალსაზრისით სასურველია აისახოს ქართული ენის ანოტირებულ კორპუსში, რისთვისაც აუცილებელია ანოტაციის დონეებში ახალი, პრაგმატიკული დონის შემოღება.

თავი მეოთხე

ქართული ენის ეროვნული კორპუსი

1. ეროვნული კორპუსის განსაზღვრებისათვის

როგორც ზემოთ უკვე აღვნიშნეთ, ენის ეროვნული კორპუსი არის სახელმწიფო ენისათვის შექმნილი ვირტუალური სივრცე, რომელშიც თავმოყრილია სახელმწიფო ენის ისტორიული და თანამედროვე მემკვიდრეობა ელექტრონული რესურსების სახით. ენის ეროვნული კორპუსი სახელმწიფო ენის სტატუსით აღჭურვილი ენისათვის შექმნილი კორპუსია, რომელიც გამოიყენება როგორც სახელმწიფო ინტერესებისათვის (სახელმწიფო ენის ნორმალიზაცია-სტანდარტიზაცია, ენის ეკოლოგიის დაცვა, ენობრივი პოლიტიკის დაგეგმვა და ა.შ.), ისე ქვეყანაში მეცნიერების სხვადასხვა დარგის, განსაკუთრებით ციფრული პუმანიტარიის განვითარებისათვის. ეროვნული კორპუსი სახელმწიფო ენაზე ზრუნვის განხორციელებისათვის ეფექტიანი და მეთოდურად გამართული საშუალებაა, რაზედაც მეტყველებს, მაგ., ინგლისური ენისათვის შექმნილი სამი კორპუსის არსებობა: ბრიტანული ეროვნული კორპუსი, ამერიკული ეროვნული კორპუსი, ავსტრალიური ეროვნული კორპუსი. მიუხედავად იმისა, რომ სამივე დასახელებული კორპუსი ინგლისური ენის რესურსების ნაკრებს წარმოადგენს, თითოეული მათგანი შესაბამისი ქვეყნის სახელმწიფო ენის (ბრიტანული ინგლისური, ამერიკული ინგლისური, ავსტრალიური ინგლისური) ამსახველი ელექტრონული რესურსების სტრუქტურირებულ მონაცემთა ბანკია.

ქართული ენის ეროვნული კორპუსი წარმოადგენს ვირტუალურ სივრცეს, რომელიც მოიცავს ქართული ენის ისტორიული არსებობის განმავლობაში შექმნილ წერილობით ძეგლებსა და ქართული ენის ზეპირმეტყველების გაციფრებულ რესურსებს და საშუალებას აძლევს მომხმარებელს, გამოიყენოს იგი ლინგვისტური და ინტერდისციპლინური – როგორც დიაქრონიული, ისე სინქრონიული – კვლევებისათვის. ქართული ენის ეროვნული კორპუსი შეიქმნა საერთაშორისო სამეცნიერო პროექტების ფარგლებში გერმანელი და ქართველი მეცნიერების თანამშრომლობის შედეგად და დიგიტალური ქართველოლოგიის განვითარების ერთ-ერთ ძირითად რესურსს წარმოადგენს.

2. ქართული ენის ეროვნული კორპუსის შექმნის ისტორია და საფუძვლები

ქართული ენის ეროვნული კორპუსის ტექნოლოგიური ჩარჩოს შექმნა დაიწყო 2012 წელს, ფოლკსვაგენის ფონდის მიერ დაფინანსებული საერთაშორისო სამეცნიერო პროექტის „ქართული ენის ეროვნული კორპუსი: ტექნოლოგიური ჩარჩოს შექმნა“ (ხელმ.: იოსტ გიპერტი და მანანა თანდაშვილი) ფარგლებში. ქართული ენის ეროვნული კორპუსის ელექტრონული რესურსების საფუძველს წარმოადგენდა ის ელექტრონული რესურსების ნაკრები, რომელიც იქმნებოდა სამეცნიერო პროექტების სახით ფრანკფურტის გოეთეს სახ. უნივერსიტეტის ემპირიული ენათმეცნიერების ინსტიტუტში გერმანელ და ქართველ მეცნიერთა მონაწილეობით და განთავსებული იყო გერმანიაში (TITUS-სა და ARMAZI-ს ქართული რესურსები), ნორვეგიაში (Clarino), ჰოლანდიასა (Max Plank Institute, Language Archive) და საქართველოში (დიალექტური კორპუსი, Opentext და lib.ge).

ამდენად, ქართული ენის ეროვნული კორპუსი ეყრდნობა დიგიტალურ რესურსებს, რომელიც გასული საუკუნის 80-იანი წლებიდან იქმნებოდა საქართველოსა და მის ფარგლებს გარეთ და მოიცავს როგორც წერილობით, ისე ზეპირმეტყველების წერილობით დოკუმენტირებულ და გაციფრებულ რესურსებს:

1. **TITUS**-ის ქართული რესურსები (გერმანია, ფრანკფურტის უნივერსიტეტი, პროექტის ხელმძღვანელი იოსტ გიპერტი), რომელიც მოიცავს ძველი და საშუალი ქართულის ორიგინალურ და ნათარგმნ ძეგლებს.
2. კავკასიური ენების ელექტრონული პორტალის **ARMAZI**-ს ქართული რესურსები (გერმანია, ფრანკფურტის უნივერსიტეტი, პროექტის ხელმძღვანელები: იოსტ გიპერტი და მანანა თანდაშვილი), რომელშიც შესულია 1999-2008 წლებში გაციფრებული რესურსები ძველი, საშუალი და ახალი ქართულიდან.
3. სამეცნიერო პროექტები **ECLinG** (საფრთხეში მყოფი კავკასიური ენები საქართველოში) და **SSGG** (სოციოლინგვისტური სიტუაცია თანამედროვე საქართველოში) ფარგლებში ჩაწერილი ქართული ენის ზეპირმეტყველების აუდიო-ვიდეო მასალები, რომელიც ინახება ჰოლანდიაში, ნეიმეხენის მაქს პლანკის ინსტიტუტის ენის დიგიტალურ არქივში (გერმანია, ფრანკფურტის უნივერსიტეტი, პროექტის ხელმძღვანელები: იოსტ გიპერტი და მანანა თანდაშვილი).
4. ქართული დიალექტური კორპუსის (**GDC**) რესურსები (საქართველო, არნ. ჩიქობავას სახ. ენათმეცნიერების ინსტიტუტი, ხელმძღვანელი: მარინე ბერიძე), რომელიც სამეცნიერო პროექტების განხორციელების შედეგად იქმნებოდა და

ფინანსდებოდა ღია საზოგადოება საქართველოსა და რუსთაველის მეცნიერების განვითარების ეროვნული ფონდის მიერ. აღნიშნული კორპუსი წარმოადგენს ქართული ენის ზეპირმეტყველების კორპუსს და მოიცავს უნიკალურ ეთნოლინგვისტურ მასალას. მასში შესულია როგორც არნ. ჩიქობავას სახ. ენათმეცნიერების არქივში დაცული დიალექტური ექსპედიციების მასალების გაციფრებული ვერსიები, ისე საქართველოში გავრცელებული თანამედროვე დიალექტური მეტყველება და ისტორიული დიასპორების ზეპირმეტყველების უნიკალური ნიმუშები.

5. **GEKKO**-ს რესურსები (ნორვეგია, პროექტის ავტორი პაულ მოირერი);
6. ქართული პრესის ელექტრონული არქივი **Opentext** (პროექტის ავტორი ირაკლი იაკობაშვილი) – პირველი პრესის კორპუსი, რომელიც შეიქმნა საქართველოში და აერთიანებს პოსტსაბჭოთა პრესის დიგიტალურ მასალებს.
7. ღია ელექტრონული რესურსი **lib.ge** (პროექტის იდეის ავტორი გიორგი კეკელიძე), თავისუფალი რესურსი, რომელშიც შესულია XIX-XX საუკუნეების ქართული ლიტერატურული და პუბლიცისტური მემკვიდრეობა.

3. ქართული ენის ეროვნული კორპუსის მოცულობა

ქართული ენის ეროვნული კორპუსი დღევანდელი მონაცემების მიხედვით 202 მილიონზე მეტი ტოკენის, ანუ სიტყვაფორმის მომცველია (შდრ. ბრიტანული ეროვნული კორპუსი – 100 მილიონი ტოკენი, სომხური ეროვნული კორპუსი – 110 მილიონი ტოკენი, რუსული ეროვნული კორპუსი – 150 მილიონი ტოკენი). მასში შესულია როგორც ქართული ენის ისტორიული მემკვიდრეობა (V-XVIII საუკუნეების ლიტერატურული და ისტორიული ძეგლები), ისე XIX-XX საუკუნეების ქართული ლიტერატურა და 90-იანი წლებიდან მოყოლებული ქართული პრესა. ამჟამად მიმდინარეობს ქართული ენის კორპუსის თემატური ქვეკორპუსებით განვრცობა. კორპუსის განვითარების მთავარ მიზანს წარმოადგენს კორპუსის ჟანრობრივი დაბალანსება, კერძოდ სამეცნიერო, იურიდიული, პოლიტიკური და სოციოლექტური მასალებით მისი გაჯერება.

4. ქართული ენის ეროვნული კორპუსის ეტაპობრივი განვითარება

ქართული ენის ეროვნული კორპუსის ტექნოლოგიური ჩარჩოს შექმნა და მისი დიგიტალური რესურსებით გაჯერება, როგორც ზემოთ უკვე აღვნიშნეთ, 2012 წელს დაიწყო. 2013 წლის 29 სექტემბერს ხელი მოეწერა ურთიერთთანამშრომლობის მემო-

რანდუმს ფრანკფურტის გოეთეს სახ. უნივერსიტეტის ემპირიული ენათმეცნიერების ინსტიტუტსა და საქართველოს პარლამენტის ეროვნულ ბიბლიოთეკას შორის, რომელიც ითვალისწინებდა ფრანკფურტის უნივერსიტეტში განვითარებული ქართული ენის ეროვნული კორპუსის საქართველოში გადმოტანას და გრძელვადიან თანამშრომლობას ქართული ენის ეროვნული კორპუსის შემდგომი განვითარების მიზნით.



მემორანდუმის ხელმოწერა ფრანკფურტის გოეთეს უნივერსიტეტსა და საქართველოს პარლამენტის ეროვნულ ბიბლიოთეკას შორის[132].
ფოტოზე: მანანა თანდაშვილი, ფრიდონ საყვარელიძე და გიორგი კეკელიძე

ქართული ენის ეროვნული კორპუსის განვითარების შესახებ თანამშრომლობა სამ ეტაპად გაიწერა:

1. **პირველ ეტაპზე** (2013-2015 წლები) შემუშავდა ქართული ენის ეროვნული კორპუსის ტექნოლოგიური ჩარჩოს კონცეფცია. კონცეფციის შესაბამისად აღნიშნულ წლებში განხორციელდა შემდეგი სამუშაოები:
   - ქართული ენისათვის არსებული ელექტრონული რესურსების ინტეგრაცია ქართული ენის ეროვნულ კორპუსში და მისი გაერთმნიშვნელიანება;
   - ანოტაციისა და კორპუსის სტრუქტურის ვალიდურობის შუალედური ევალუაცია;
   - პარალელური კორპუსებისა და თარგმნილი ტექსტების ჯვარედინი რეფერენციის სისტემების განვითარება;
   - საძიებო სისტემის KRIC-პრინციპების შემუშავება და აპრობაცია;
   - საძიებო სისტემის განვითარება სინქრონიული და დიაქრონიული კვლევების განსახორციელებლად;
   - კორპუსის ანოტირების საშუალებების (tools) განვითარება
   - ტექსტების ანოტირება:

[132] იხ. https://www.youtube.com/watch?v=vZB3tQKs9_o

ა. ტექსტების მეტამონაცემებით აღჭურვა – კორპუსის შემადგენელ თითოეულ ტექსტზე საერთაშორისო სტანდარტების (TEI – Text Encoding Initiative) მიხედვით განსაზღვრული მეტამონაცემების მიმაგრება კორპუსში ინტერდისციპლინარული კვლევის განხორციელების მიზნით;

ბ. ტექსტის მარკირების საგანგებო ნიშნულების სისტემით აღჭურვა (textual markup);

გ. ტექსტების ლინგვისტური ანოტაციის პრინციპების შემუშავება ანოტაციის დონეების მიხედვით;

დ. გრამატიკულ მახასიათებელთა სისტემატიზაცია;

2. მეორე ეტაპზე (2015-2019 წლები) შემუშავდა ოქროს სტანდარტის (პროტოტიპული) კორპუსის კონცეფცია, კერძოდ განხორციელდა:
   - ოქროს სტანდარტის ვალიდურობის შუალედური ევალუაცია და სტანდარტის დახვეწა;
   - ქართული ენის ეროვნული კორპუსის დაბალანსების სტრატეგიის შემუშავება და თემატური განვრცობა (კორპუსი განიერცო იურიდიული და პოლიტიკური ქვეკორპუსებით);
   - ქართული ენის ეროვნული კორპუსის რეფერენციალური შეპირისპირება პარალელურ კორპუსებთან;
   - კორპუსის განვრცობისა და დაბალანსების საშუალებების (tools) შემდგომი განვითარება;
   - ქართული ენის ეროვნული კორპუსის, როგორც ღია კორპუსის, განვითარების ნორმების დადგენა;
   - ოქროს სტანდარტის შემდგომი დამუშავება:
     ა. ოქროს სტანდარტის კორპუსის მოცულობის განსაზღვრა,
     ბ. ჟანრობრივი მრავალფეროვნების განსაზღვრა ოქროს სტანდარტის კორპუსში,
   - თეისებრიეი მახასიათებლების დადგენა და კორპუსის დამუშავების ხარისხის განსაზღვრა,
   - დისამბიგვირება;
3. მესამე ეტაპზე (2020 წლიდან) დაგეგმილია კორპუსის შემდგომი დაბალანსება. კერძოდ:
   - ოქროს სტანდარტის ვალიდურობის საბოლოო ევალუაცია;
   - ოქროს სტანდარტის დადგენილი ნორმების შეფასება მთლიან კორპუსზე;

- კორპუსის დადგენილი ნორმების მიხედვით დაბალანსება;
- კორპუსის დაბალანსების შუალედური ევალუაცია;
- კორპუსის შემდგომი განვრცობის ნორმალიზაცია;
- ქართული ენის ეროვნული კორპუსის, როგორც ღია კორპუსის, ექსპლუატაცია.

ქართული ენის ეროვნული კორპუსის კონცეფცია და განვითარების გეგმა შედგენილია კორპუსის დირექტორთა საბჭოს მიერ, რომელშიც შედიან: პროფ. იოსტ გიპერტი (ფრანკფურტი, გერმანია), პროფ. მანანა თანდაშვილი (ფრანკფურტი, გერმანია), დოქტ. პაულ მოირერი (ბერგენი, ნორვეგია), პროფ. მარინე ბერიძე (თბილისი, საქართველო), ირაკლი ღარიბაშვილი (თბილისი, საქართველო), გიორგი კეკელიძე (თბილისი, საქართველო).

დღეისათვის ქართული ენის ეროვნული კორპუსის სარკისებური სერვერი განთავსებულია საქართველოს პარლამენტის ეროვნულ ბიბლიოთეკაში მისამართზე: www.gnc.gov.ge.

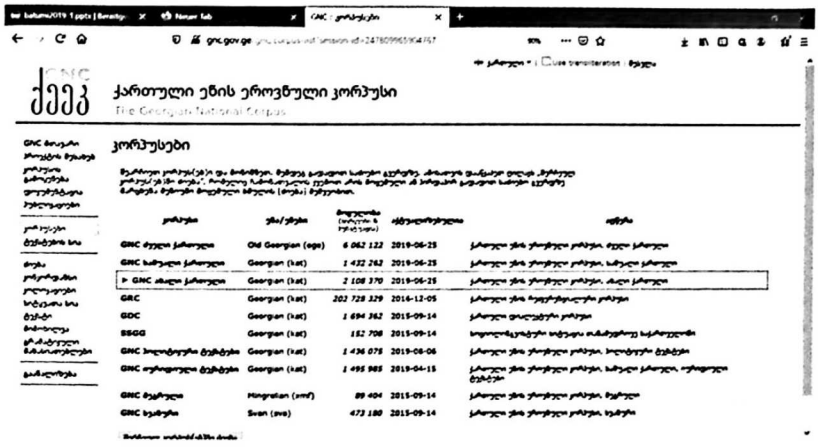
5. ქართული ენის ეროვნული კორპუსის გამოყენება

ქართული ენის ეროვნული კორპუსი ღია რესურსია და მისი გამოყენება შეუძლია ნებისმიერ მსურველს. ამ ქვეთავში აღწერილია კორპუსის გამოყენების წესები და შესაძლებლობები.

კორპუსის ეფექტიანი გამოყენების უზრუნველსაყოფად, უპირველეს ყოვლისა, საჭიროა განისაზღვროს საძიებო სივრცე. ეს სივრცე შეიძლება იყოს ქრონოლოგიური და/ან თემატური ნიშნით შერჩეული ქვეკორპუს(ებ)ი.

5.1. (ქვე)კორპუს(ებ)ის გამოყენება

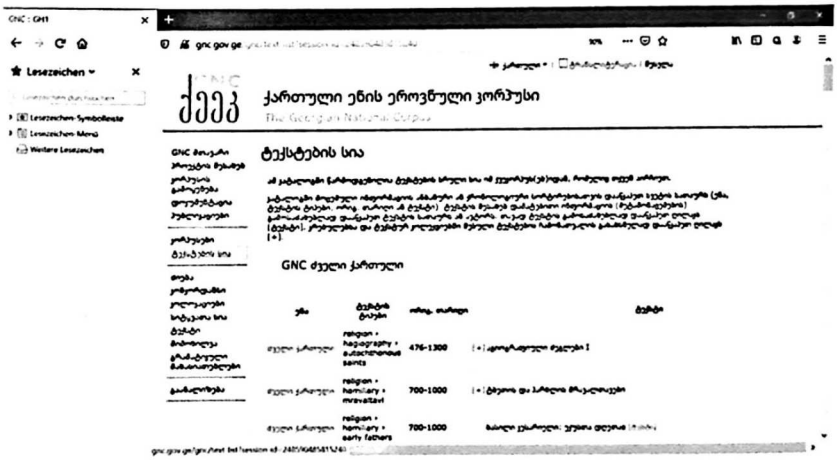
ქართული ენის ეროვნულ კორპუსში განთავსებული ყველა ქვეკორპუსი ჩამოთვლილია **კორპუსების სიაში**. თქვენ შეგიძლიათ აირჩიოთ ერთი ქვეკორპუსი ან ერთდროულად რამდენიმე ქვეკორპუსი. მაგალითად, ძიება განახორციელოთ ქართული ენის განვითარების რომელიმე პერიოდის ფარგლებში და აირჩიოთ რომელიმე ქვეკორპუსი ჩამონათვალიდან: ძველი ქართული ენის კორპუსი, საშუალო ქართული ენის კორპუსი, ახალი ქართული ენის კორპუსი ან ქართული ენის რეფერენციალური კორპუსი, რომელიც სამსავე აქ დასახელებულ კორპუსებს მიემართება და დიაქრონიული თვალსაზრისით ძიების საშუალებას გვთავაზობს. ქვეკ-ში შესაძლებელია ძიება ასევე ცალკეული პროექტების ფარგლებში: ECLinG და SSGG.



დააწკაპუნეთ თქვენ მიერ ამორჩეულ ქვეკორპუსს. მონიშნული კორპუს(ებ)ი ასეთ შემთხვევაში გააქტიურდება შემდგომი ძიების ან სხვა სახის კორპუსული ოპერაციების ჩასატარებლად.

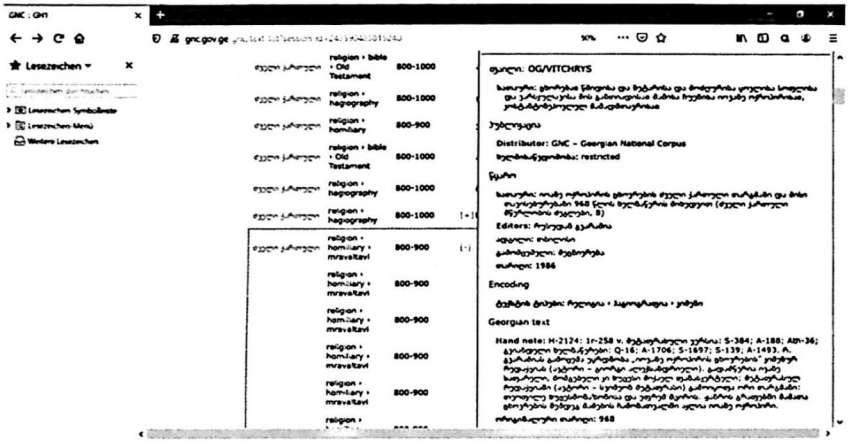
5.2. კორპუსის ტექსტების გამოტანა/ჩვენება

ქეეკ-ში შემავალი ქვეკორპუსების ტექსტები, როგორც ძველი ქართულის (GNC ძველი ქართული), ისე საშუალი ქართულის (GNC საშუალი ქართული) და ახალი ქართული ენისათვის (GNC ახალი ქართული), ჩამოთვლილია ტექსტების სიაში. თითოეულ ტექსტს თან ახლავს მეტამონაცემი ავტორის, სათაურის, შექმნის თარიღის, ენობრივი ვარიანტის და ტექსტის ჟანრის შესახებ.

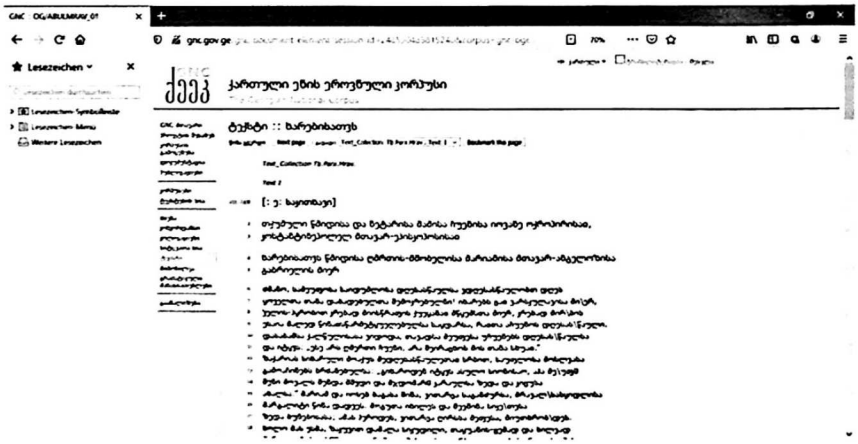


კრებულებსა და ტექსტურ კოლექციებში შესული ტექსტების ჩამონათვალის გასახსნელად დააწკაპუნეთ ნიშანს [+], რომელიც განთავსებულია კრებულის ან კოლექციის სათაურის ან ავტორის სახელის მარცხენა მხარეს.

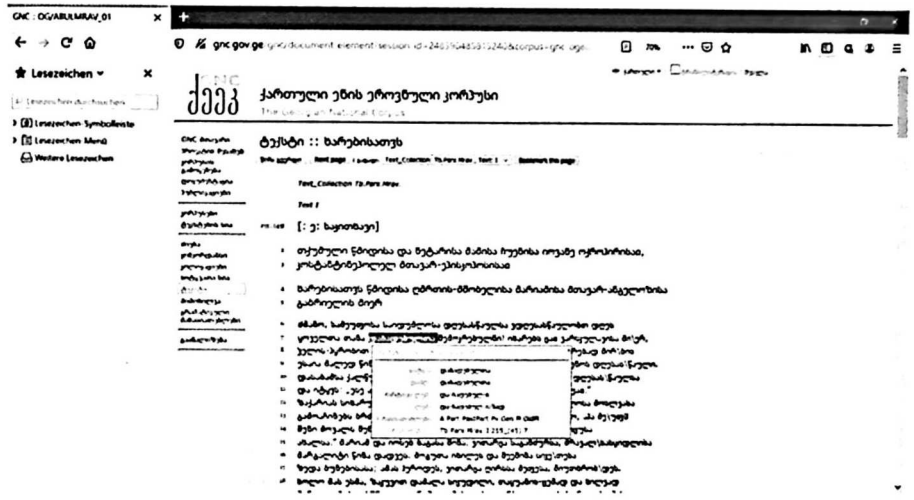
თუ თქვენ დააწკაპუნებთ ტექსტის სათაურს, გამოჩნდება მეტამონაცემები მონიშნული ტექსტის შესახებ (ტექსტის წყაროს, გამოცემის, გამოყენების შესაძლებლობის შესახებ და სხვა).



თავად ტექსტის გასახსნელად დააწკაპუნეთ ღილაკს [ტექსტი], რომელიც განთავსებულია სათაურის მარჯვენა მხარეს. შერჩეული ტექსტი გამოჩნდება ახალ გვერდზე. აქედან თქვენ შეგიძლიათ გადახვიდეთ მომდევნო გვერდებზე, თავებზე და დაბრუნდეთ უკან.

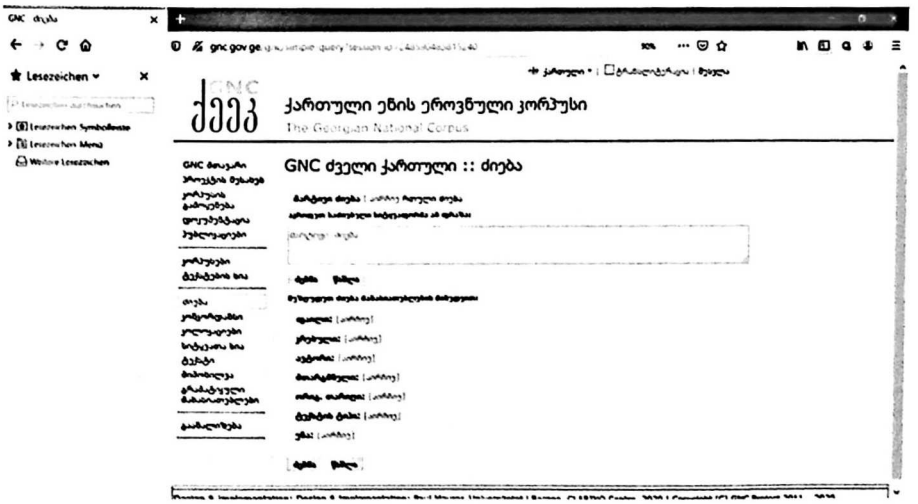


თუ თქვენ დააწკაპუნებთ ტექსტში მოცემულ რომელიმე სიტყვას, გამოჩნდება ფანჯარა, რომელშიც მოცემულია მონიშნული სიტყვის გრამატიკული მახასიათებლები (სიტყვის ნორმირებული ფორმა, ლემა და მორფოსინტაქსური ნიშნები).



5.3. კორპუსში ძიება

ამორჩეულ (ქვე)კორპუსში ან (ქვე)კორპუსებში ძიების განსახორციელებლად მთავარი გვერდის მენიუში აირჩიეთ ღილაკი ძიება.

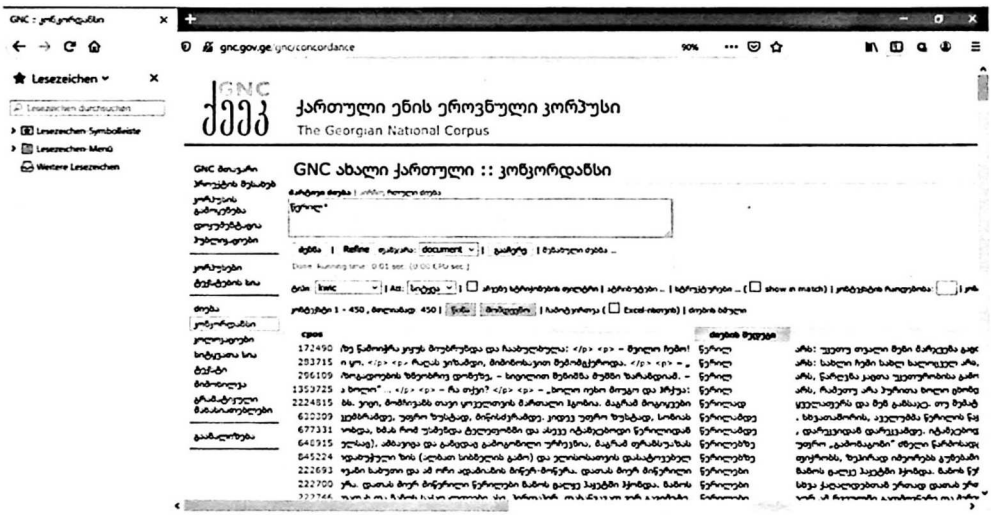


ქართული ენის ეროვნული კორპუსის სამომხმარებლო ნიღაბი კორპუსში ძებნის სამგვარ შესაძლებლობას (მოდუსს) გთავაზობთ.

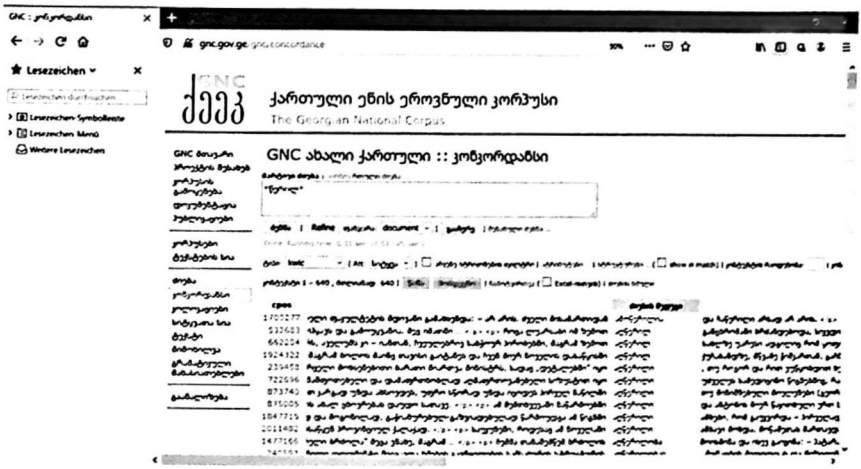
5.4. საბაზისო ძებნა

ძებნის საბაზისო მოდუსის არჩევის შემთხვევაში თქვენ შეგიძლიათ საძიებო ველში შეიტანოთ სიტყვა, ფრაზა ან წინადადება. აკრიფეთ თქვენთვის სასურველი ფორმა და დააწკაპუნეთ ღილაკს ძებნა. თქვენ არ გჭირდებათ ბრჭყალების, საგანგებო ნიშნების ან რაიმე სხვა სახის მარკირების გამოყენება. აკრიფეთ სიტყვები ისე, როგორც ისინი იწერება. საჭიროების შემთხვევაში, შეგიძლიათ გამოიყენოთ ნიშანი (*), რომელიც ძებნის პროცესში შეესაბამება „ნებისმიერ ნიშანს".

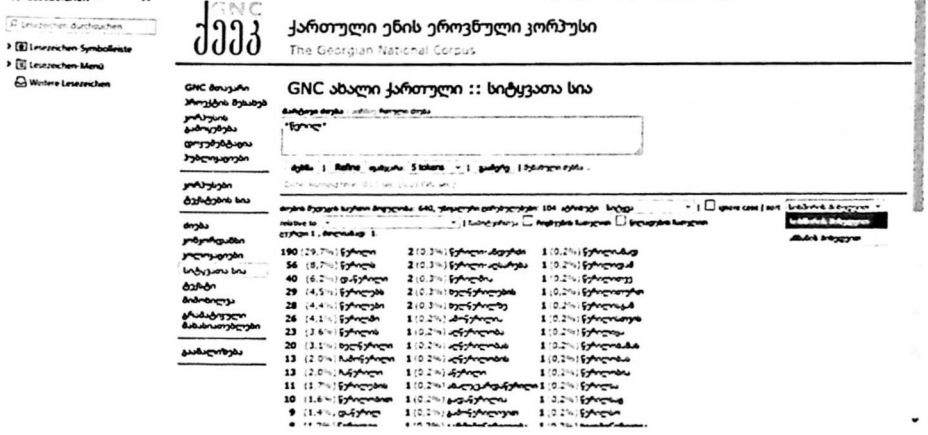
მაგალითად: თუ თქვენ საძიებო ველში ჩაწერთ „წერილ*", მაშინ კორპუსში მოიძებნება არა მარტო ისეთი ფორმები, როგორიცაა *„წერილი", „წერილს", „წერილებს"*, არამედ თანდებულიანი ფორმებიც, მაგალითად, „წერილში", „წერილიდან" და ა.შ.



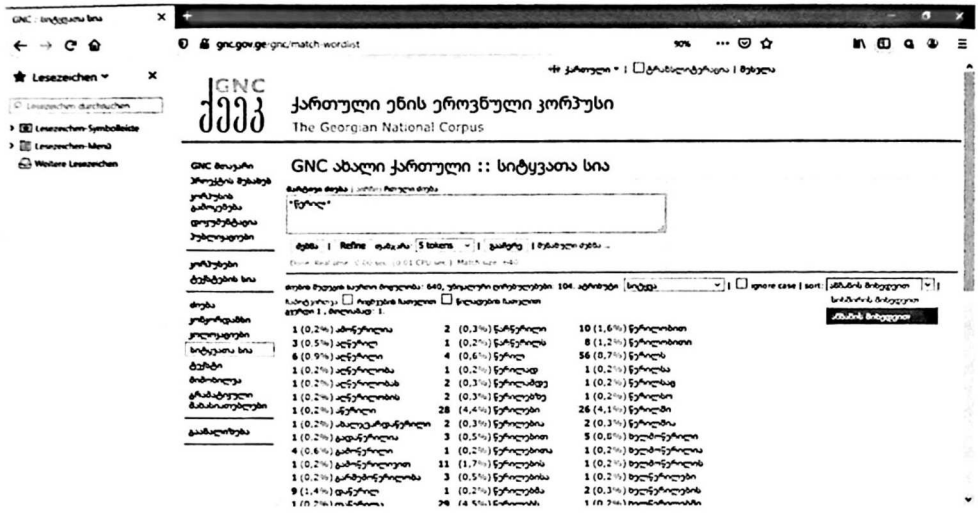
თუ თქვენ გამოიყენებთ ნებისმიერობის ნიშანს (*) როგორც ფუძის თავში, ისე ბოლოში, მაშინ კორპუსში მოიძებნება ყველა სიტყვაფორმა, რომელიც შეიცავს ფუძეს *„წერილ": ამოწერილია, აღწერილ, აღწერილობა, წერილობით, გადაწერილ, ჩასაწერ, წერილებს, წერილობითი, ხელწერილი, დაწერილმა, ჩაწერილ, დაწერილ, მოწერილ, მიწერილია* და ა.შ.



თუ თქვენ დააწკაპუნებთ ღილაკს „სიტყვათა სია“, ავტომატურად გენერირდება იმ სიტყვათა ჩამონათვალი, რომელიც შეიცავს აღნიშნულ ფუძეს. ამასთან, სიტყვათა ჩამონათვალი შეგიძლიათ ავტომატურ რეჟიმში დაალაგოთ სიხშირის მიხედვით, ან ანბანის მიხედვით.

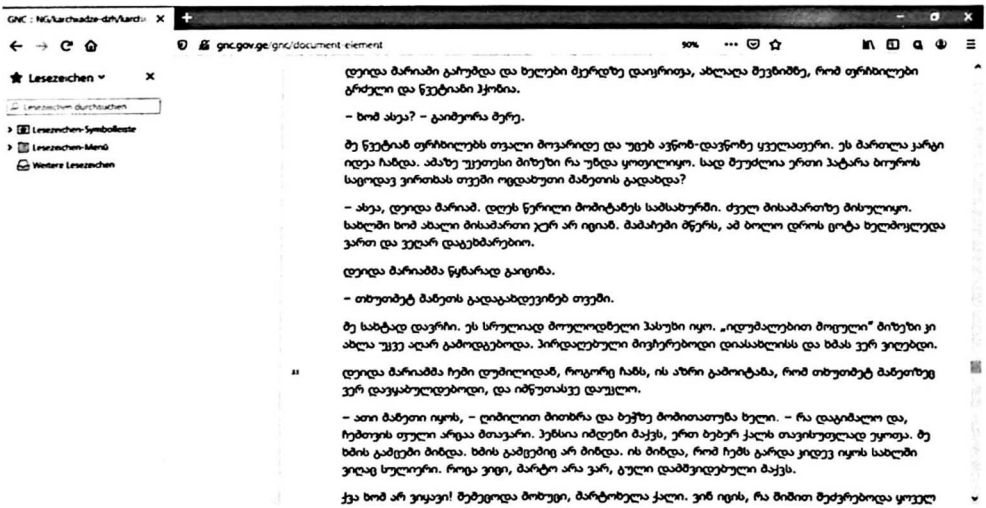


სიტყვათა სია დალაგებული სიხშირის მიხედვით



სიტყვათა სია დალაგებული ანბანის მიხედვით

საჭიროების შემთხვევაში, სიტყვათა სიიდან შეგიძლიათ გადახვიდეთ კონტექსტში. ამისათვის უნდა დააწკაპუნოთ სიტყვათა ჩამონათვალში თქვენთვის სასურველ სიტყვას.



დეიდა მარიამი გაიხმდა და ხელები მკერდზე დაიკრიფა, ახლაღა შევნიშნე, რომ თრიმლები გრძელი და ნვეტიანი ჰქონია.

– ხომ ასეა? – გაიმეორა მერე.

მე ნვეტიან თრიმლებს თვალი მოვარიდე და უცებ ავწონ-დავწონე ყველაფერი. ეს მართლა კარგი იდეა ჩანდა. ამაზე უკეთესი მიზეზი რა უნდა ყოფილიყო. სად მეუძლია ერთი პატარა ბიჭის საფიდან ვირთხას თევში ოფლახუნი მაბეთის გადახდა?

– ასეა, დეიდა მარიამ. დღეს ნერილი მომიტანეს სამსახურში. ძველ მისამართზე მისულიყო. სახლში ხომ ახალი მისამართი ჯერ არ იციან. მამაჩემი მწერს, ამ ბოლო დროს ცოტა ხელმოკლედ ვართ და ვეღარ დაგეხმარებით.

დეიდა მარიამმა წყნარად გაიცინა.

– თხუთმეტ მანეთს გადაგახდევინებ თევში.

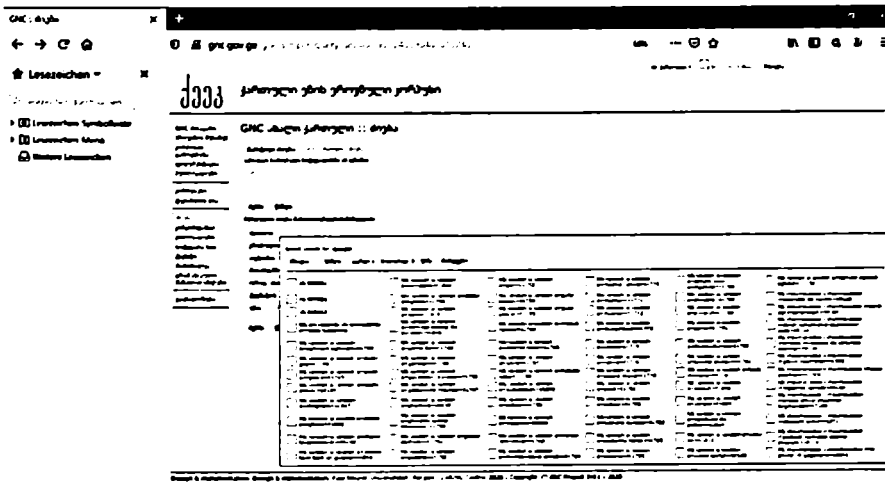
მე სახტად დავრჩი. ეს სრულიად მოულოდნელი პასუხი იყო. „იდუმალებით მოცული" მიზეზი კი ახლა უკვე აღარ გამოდგებოდა. პირდაღებული მივჩერებოდი დისახლისს და ხმას ვერ ვიღებდი.

დეიდა მარიამმა ჩემი დუმილიდან, როგორც ჩანს, ის აზრი გამოიტანა, რომ თხუთმეტ მანეთზე ვერ დავყაბულდებოდი, და იმწუთასვე დაუკლო.

– ათი მანეთი იყოს, – ღიმილით მითხრა და ბუჭზე მომითათუნა ხელი. – რა დაგიმალო და, ჩემთვის თვალი არცაა მთავარი. ეუნსია იმდენი მაქვს, ერთ ბებერ ქალს თავისუფლად უყოფა. მე ხმის გამცემი მიმდა. ხმის გამცემიც არ მიმდა. ის მიმდა, რომ ჩემს გარდა კიდევ იყოს სახლში ვიღაც სულიერი. როგა ვიცი, მარტო არა ვარ, გული დამშვიდებული მაქვს.

ჭკა ხომ არ ვიყავი! მომეგონა მოხუცი, მარტოხელა ქალი. ვინ იცის, რა მიმით მეძერებოდა ყოველ

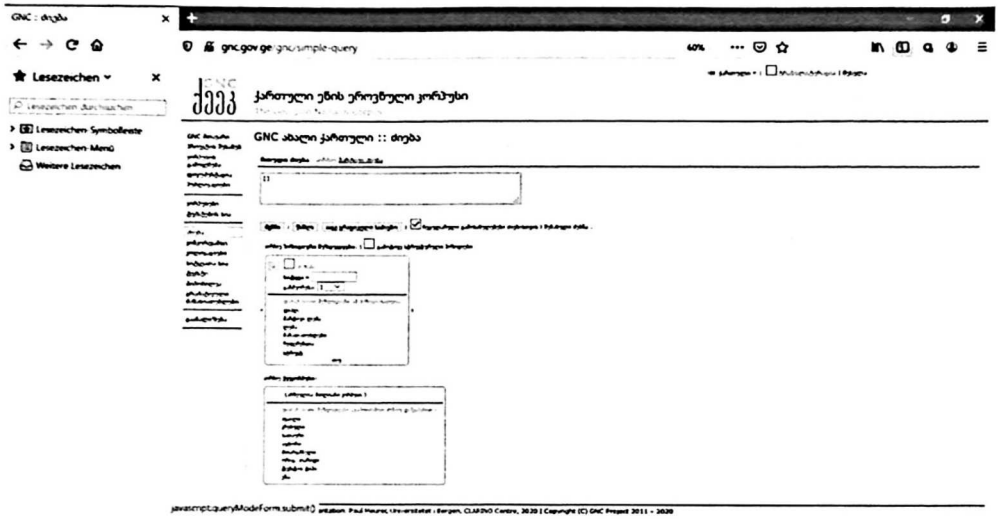
თქვენ შეგიძლიათ ძებნა განახორციელოთ მთლიან კორპუსში ან შეზღუდოთ იგი ქვეკორპუსის ფარგლებში, მაგალითად, ენის რომელიმე პერიოდის (ძველი ქართული, საშუალი ქართული, ახალი ქართული) ან ჟანრ(ებ)ის, ავტორ(ებ)ის, ნაწარმოების მიხედვით. დააწკაპუნეთ ღილაკს არჩევა იმ ატრიბუტების გვერდით, რომელთა მიხედვითაც გსურთ ძიების შეზღუდვა. ასეთ შემთხვევაში გაიხსნება ფანჯარა, რომელშიც შეგიძლიათ მონიშნოთ თქვენთვის სასურველი პერიოდი, ჟანრი, ავტორი, ნაწარმოები ან სხვ.



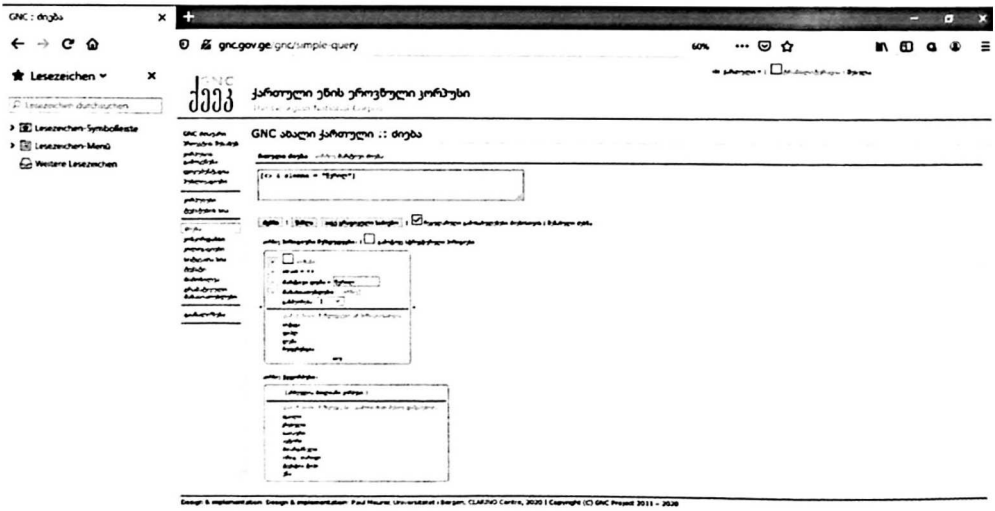
5.5. გაფართოებული ძებნა

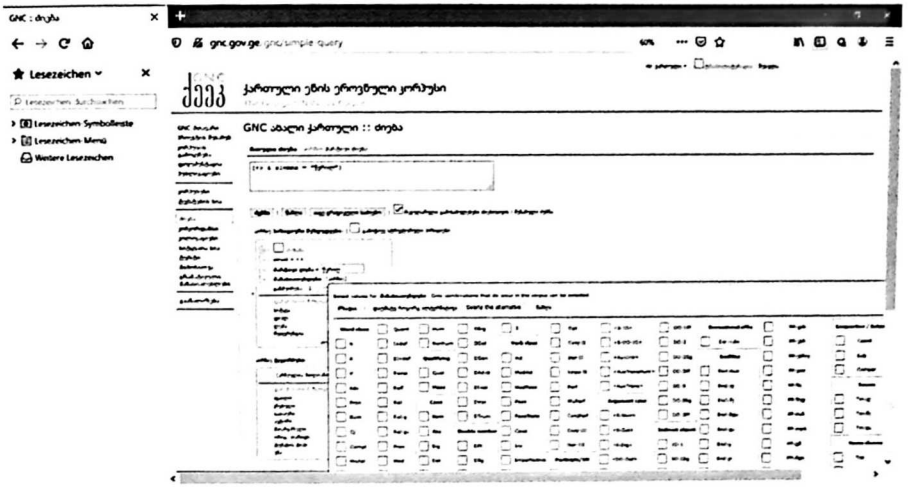
საბაზისო მოდულის ფარგლებში ძებნა ყოველთვის ვერ აკმაყოფილებს ჩვენს სამეცნიერო ინტერესებს. ბევრი საინტერესო საკითხი საგანგებო, გაფართოებულ ძებნას მოითხოვს. გაფართოებული ძებნის მოდულის ფარგლებში თქვენ შესაძლებლობა გეძლევათ, ძებნა განახორციელოთ რეგულარული გამოსახულებების გათვალისწინებით. რეგულარული გამოსახულებები (Regular expression; შემოკლებით **regex, regexp**) კომპიუტერულ მეცნიერებაში გამოყენებული საშუალებაა ერთი ჩანაწერით იპოვოს სასურველი ბლოკი ტექსტში. გაფართოებული ძებნის დროს გამოსახულება შეიძლება ძლიერ გართულდეს და კომპლექსური ხასიათი შეიძინოს. მრავალი პროგრამირების ენა იყენებს ან უზრუნველყოფს კორპუსში რეგულარული გამოსახულებით ძიებას.

თქვენ შეგიძლიათ კორპუსში ძებნა განახორციელოთ მარტივი ლემის საშუალებით:

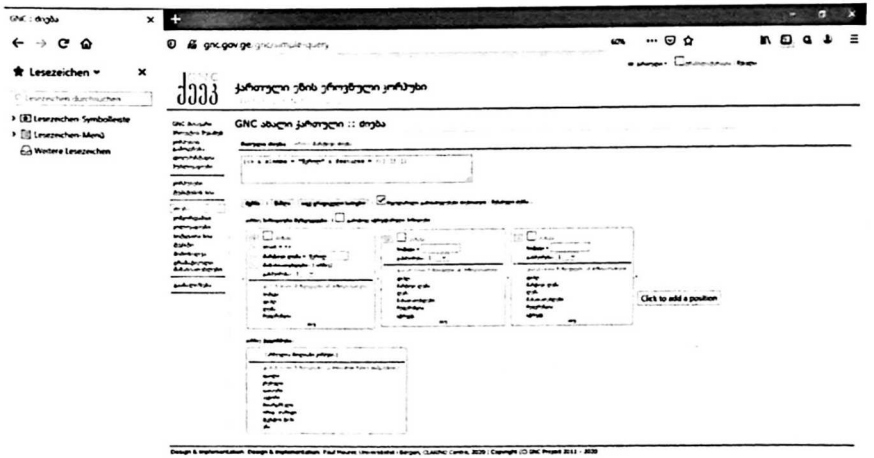


დამატებით შესაძლებელია ძებნის განხორციელება მახასიათებლების მიხედვით. ამისათვის გააააქტიურეთ ღილაკის „მახასიათებლები“ გვერდით განთავსებული ღილაკი „აირჩიე“:





ამასთან, ძებნის ეს მოდუსი საშუალებას გაძლევთ საგანგებოდ შემუშავებული მენიუს გამოყენებით კომბინირებულად, რამდენიმე ნიშნის მიხედვით განახორციელოთ ძებნა ან გარკვეული პირობების გათვალისწინებით შეზღუდოთ (სიტყვების, სტრუქტურული ელემენტების) ძებნის პროცესი კორპუსში. ამისათვის დააწკაპუნეთ ნიშანს + და ახლად გამონათებულ ფანჯარაში შეიტანეთ თქვენთვის სასურველი ლემა ან ლემის მახასიათებლები:



თქვენ ამ შემთხვევაშიც შეგიძლიათ დამატებით გამოიყენოთ იგივე მენიუ, რომელიც მარტივი ძებნის დროს გამოიყენება, ანუ განახორციელოთ ძებნა ენის რომელიმე პერიოდის (ძველი ქართული, საშუალი ქართული, ახალი ქართული) ან ჟანრ(ებ)ის, ავტორ(ებ)ის, ნაწარმოების ფარგლებში.

5.6. CQL-ის გამოყენება ძებნისას

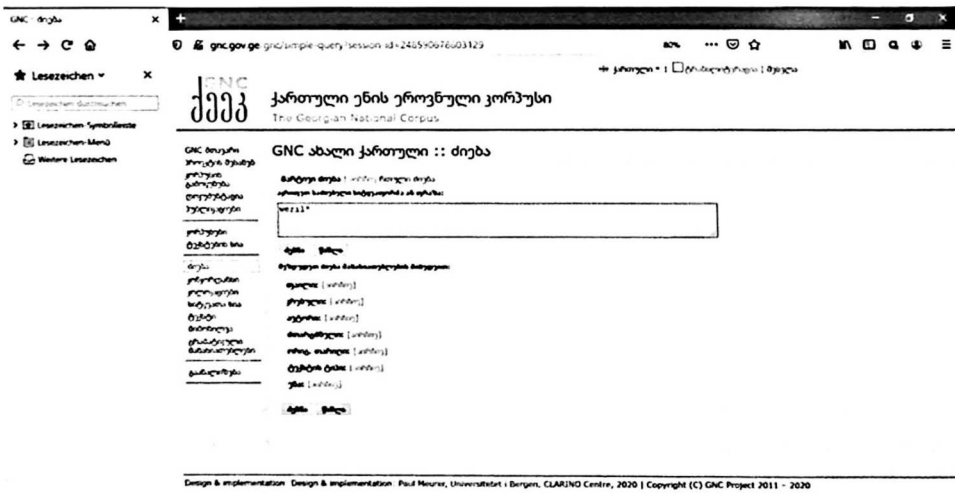
კორპუსის საძიებო სისტემასთან მიმართების პირდაპირი და ყველაზე მოსახერხებელი საშუალება არის ძიების განხორციელება CQL-ის (Corpuscle Query Language) მეშვეობით. კორპუსში ძიების საგანგებო ენა აღწერილია საგანგებო ბმულში **ძიების ენა** (იხ. ინგლისურად).

კორპუსში მარტივი ან გაფართოებული ძებნა შეგიძლიათ განახორციელოთ რეგულარული გამოსახულებების მიხედვითაც, თუ თქვენ დააწკაპუნებთ ღილაკს **საძიებო გამოსახულებები**.

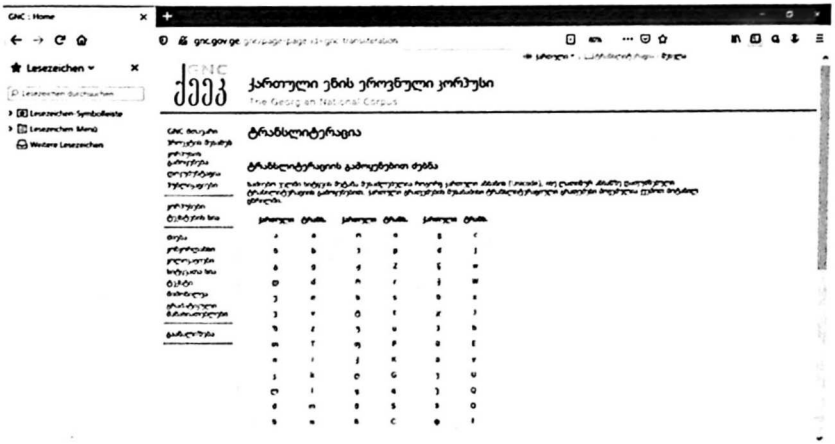
თუ თქვენ გსურთ შესაბამის ველში გრაფიკულად თავად შეიტანოთ რეგულარული გამოსახულებები, **CQL**-ძიების ენა ავტომატურად გააქტიურდება, რათა მოახდინოს თქვენი არჩევანის რეფლექტირება. ეს არის ინტუიტიური გზა იმისათვის, რომ ვისწავლოთ, როგორ განვახორციელოთ კორპუსთან მიმართებაში შეკითხვის ფორმულირება. ამ გზით თქვენ შეგიძლიათ მოახდინოთ რეგულარული გამოსახულებების მოდიფიცირება და ღილაკზე **გრაფიკული ძებნის შექმნა** დაწკაპებით განახორციელოთ განახლებული ძებნა.

5.7. ძებნა ტრანსლიტერაციის გამოყენებით

თქვენ შეგიძლიათ ძებნა განახორციელოთ როგორც ქართული ანბანით, ისე ტრანსლიტერაციის მეშვეობით ძებნის სამივე მოდუსში. ტრანსლიტერაციის შემთხვევაში შეგიძლიათ გამოიყენოთ ლათინური გრაფემები (ASCII).



ტრანსლიტერაციის წესები მოცემულია ვებგვერდზე: http://gnc.gov.ge/gnc/page?page-id=gnc-transliteration.



ტრანსლიტერაციის გამოყენებით ძებნა გამოსადეგია იმ მეცნიერთათვის, რომლებიც არ იყენებენ ქართულ ფონტებს ან იმ შემთხვევისათვის, როდესაც ქართული კლავიატურის გამოყენება მოცემულ მომენტში რაიმე მიზეზის გამო შეუძლებელია.

5.8. კონკორდანსი

საძიებო გვერდზე ძებნის განხორციელების შემდეგ შესაძლებელია კონკორდანსის გამოძახება. კონკორდანსის გენერირებისათვის დააწკაპუნეთ ღილაკს **კონკორდანსი**. ამასთან თქვენ შეგიძლიათ განსაზღვროთ კონტექსტის რაოდენობა ან კონტექსტის სიგანე.



6. სტატისტიკა ლინგვისტიკის სამსახურში

სტატისტიკის გამოყენება მეცნიერების სხვადასხვა დარგებში ფუნდამენტური კვლევების განუყოფელი ნაწილია და ამ თვალსაზრისით არც ლინგვისტიკაა გამონაკლისი, თუმცა დიგიტალური ჰუმანიტარიის განვითარებამ – ანუ დიგიტალური რესურსების დამუშავების შესაძლებლობამ დიგიტალური მეთოდების გამოყენებით – სრულიად ახალი ფუნქცია მიანიჭა კვლევის ამ მეთოდს – დიგიტალური კვლევის მეთოდებს შეემატა ახალი – **ტექსტის კვანტიტატიური კვლევის მეთოდი**, რომელიც დროთა განმავლობაში იმდენად განვითარდა, დაიხვეწა და გართულდა, რომ კვლევის მეთოდიდან ახალ მიმართულებად გადაიქცა. ტექსტის რაოდენობრივი ანალიზი არის ტექსტის დამუშავების თანამედროვე მიდგომა, რომელიც საშუალებას გვაძლევს ტექსტში „აღმოვაჩინოთ" ლატენტური (ფარული) სტრუქტურები და პროცესები, რომელთა დანახვა შეუძლებელია უშუალოდ ტექსტის „ზედაპირზე" ტექსტის გრამატიკული ან ლექსიკური ფორმების ანალიზის გზით.

ტექსტის სტატისტიკურად დამუშავების მეთოდი, ზოგადად, ეფექტურია დიდ მონაცემთა ბაზების სტატისტიკური დამუშავების შემთხვევაში, ვინაიდან დიდ მონაცემთა მასივები, რეპრეზენტაციულობიდან გამომდინარე, გამორიცხავს „ინდივიდუალურობის ფაქტორს", რომლითაც, როგორც წესი, მცირე მოცულობის ტექსტები ხასიათდება (განსაკუთრებით ლიტერატურული ტექსტები). **ტექსტის რაოდენობრივი კვლევის შემთხვევაში** კი პირიქით, კვლევა მიზნად ისახავს ტექსტის როგორც ზოგადი, ისე ინდივიდუალური მახასიათებლების გამოვლენას[133].

ტექსტის რაოდენობრივი ანალიზი გულისხმობს:

1. ტექსტების დახასიათებას პარამეტრების მეშვეობით;
2. ტექსტების შედარებას მათი მახასიათებლების მიხედვით და ამის საფუძველზე ტექსტების შემდგომ კლასიფიკაციას;
3. იმ კანონების კვლევას, რომლებიც განსაზღვრავენ ტექსტების აგების პრინციპებს და ამის საფუძველზე "ტექსტების თეორიის" შემუშავებას[134].

ტექსტი ყოველთვის შეიძლება წარმოდგენილ იქნეს როგორც ტექსტის შემადგენელი ერთეულების (entity) თანმიმდევრობა, სადაც თითოეულ ერთეულს,

---

[133] Gabriel Altmann, *Wiederholungen in Texten*. Brockmeyer, Bochum, 1988, 3.

[134] Arne Ziegler, Karl-Heinz Best, Gabriel Altmann: *A contribution to text spectra*, in: *Glottometrics* 1, 2001, S. 97–108.
https://www.ram-verlag.eu/wp-content/uploads/2018/08/g1zeit.pdf#page=100

მაგალითად, სიტყვას, აქეს დამოუკიდებელი X ცვლადის ფუნქცია. დამოკიდებული Y ცვლადის ფუნქცია შეიძლება მიენიჭოს შერჩეული ცვლადების რაოდენობას x პოზიციამდე. დამოკიდებული ცვლადი ასეთ შემთხვევაში წარმოადგენს მუდმივად მზარდ თანმიმდევრობას, რომელიც, როგორც წესი, აღჭურვილია უწყვეტი ფუნქციით.

ტექსტის შემადგენელ ერთეულად შეიძლება შერჩეულ იქნეს სიტყვათა რომელიმე კლასი, მაგალითად, მეტყველების ნაწილი ან მისი ქვეკლასი. პირველ შემთხვევაში შეგვიძლია გამოვითვალოთ არსებითი სახელის, როგორც X ცვლადის ფუნქცია, მაგალითად, ნაცვალსახელებთან შედარებით, მეორე შემთხვევაში კი არსებითი სახელის ქვეკლასები როგორც შინაარსობრივი – აბსტრაქტული, სპეციფიკური, საკუთარი და ა.შ., ისე ფორმალური თვალსაზრისით, მაგალითად, აბრევიატურები, უცხო სიტყვები და ა.შ.[135]

ტექსტის რაოდენობრივი კვლევის ეს მიდგომა ერთ-ერთ უახლეს მიდგომას წარმოადგენს კორპუსლინგვისტიკაში და წარმატებით გამოიყენება დიგიტალურ ჰუმანიტარიაში. ქვემოთ შევეცდებით წარმოვადგინოთ აღნიშნული მიდგომა ქართული ენის ეროვნული კორპუსის მაგალითზე.

ჩვენ ზემოთ უკვე ვისაუბრეთ კორპუსლინგვისტიკაში ორი უმნიშვნელოვანესი მეთოდის შესახებ: კორპუსზე ორიენტირებული მეთოდი და კორპუსით განპირობებული მეთოდი. ამათგან პირველი – **კორპუსზე ორიენტირებული მეთოდი** – „მთლიანის“ შემადგენელ ელემენტთა რაოდენობრივ მიმართებებს წარმოადგენს მზარდი ან კლებადი პრინციპით, მეორე – **კორპუსით განპირობებული მეთოდი** კი – დადგენილი რაოდენობრივი მიმართებების თვისობრივი ბუნების ახსნას „გვთავაზობს“. ქვემოთ შევეცდებით წარმოვაჩინოთ სტატისტიკური მეთოდის გამოყენების ეფექტიანობა.

ქართული ენის ეროვნული კორპუსი საშუალებას იძლევა სტატისტიკურად გამოვიანგარიშოთ სიტყვათა გამოყენების სიხშირული მაჩვენებელი. ვინაიდან ქართული ენის განვითარების (საყოველთაოდ მიღებული) პერიოდიზაცია სამ ეტაპს მოიცავს (ძველ, საშუალ და ახალ ქართულს), რაც ქართული ენის ეროვნულ კორპუსში კორპუსების ჩამონათვალში შესაბამისად არის ასახული, კორპუსი საშუალებას გვაძლევს დიაქრონიულად (რეფერენციული კორპუსის ბაზაზე) ან ცალ-ცალკე – ძველი, საშუალი ან ახალი ქართულის ბაზაზე – ვიკვლიოთ სიტყვათა სიხშირე.

---

[135] Arne Ziegler, Karl-Heinz Best, Gabriel Altmann: *A contribution to text spectra*, in: *Glottometrics* 1, 2001, S. 98.

ამისათვის კორპუსების ჩამონათვალში უნდა ავირჩიოთ ერთ-ერთი, მაგალითად, **ძველი ქართული** და მარცხენა პანელში, მენიუს ჩამონათვალში ავირჩიოთ „მიმოხილვა“. ეკრანზე გამოჩნდება ახალი – **მიმოხილვის ფანჯარა**, რომელიც გვაძლევს პირველად სტატისტიკურ მონაცემებს სხვადასხვა პარამეტრების მიხედვით.



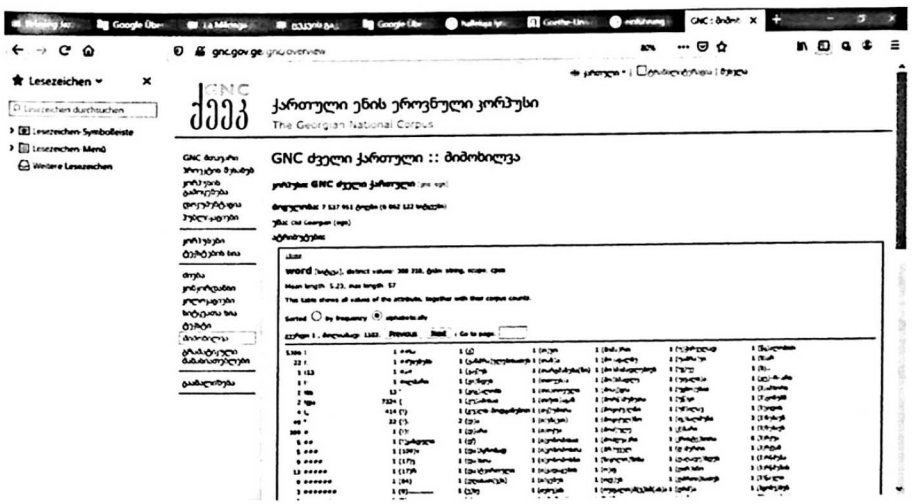
ესენია: მოცულობა (გაზომილი ტოკენებში), ენა (ჩვენ შემთხვევაში: ძველი ქართული) და ატრიბუტები, როგორიცაა სიტყვათა რაოდენობა, ლემების რაოდენობა, მახასიათებლები, ფაილების (ანუ ტექსტების) რაოდენობა, ავტორთა რაოდენობა, ტექსტის ჟანრების რაოდენობა და ა.შ.

თქვენ შეგიძლიათ აირჩიოთ რომელიმე, მაგალითად მათგანი, მაგალითად, **სიტყვათა რაოდენობა**. შედეგად ეკრანზე გამოვა ახალი ფანჯარა, რომელიც მიუთითებს არჩეულ კორპუსზე (მოცემულ შემთხვევაში **ქართული ენის ეროვნული კორპუსის ქვეკორპუსი – ძველი ქართული**), ქართული ენის პერიოდზე (მოცემულ შემთხვევაში: **ძველი ქართული**), არჩეული კორპუსის მოცულობაზე ტოკენებში (7 537 951) და სიტყვებში (6 062 122). ატრიბუტების ჩამონათვალში მითითებულია სორტირების ერთეული – სიტყვა (word) და სორტირებული სიტყვების დალაგების შესაძლებლობები, მაგალითად, სორტირებული სიტყვების დალაგება სიხშირის მიხედვით ან ანბანური წესის გამოყენებით.



სორტირება სიხშირული მაჩვენებლის მიხედვით

მენიუში თქვენ შეგიძლით აირჩიოთ სიტყვათა დალაგების პრინციპი: სიხშირული ან ანბანური. სიხშირული პრინციპით დალაგების შემთხვევაში სიტყვები დალაგდება კლებადობის პრინციპით (ზუსტი რაოდენობის თანხლებით), ხოლო ანბანური პრინციპის შემთხვევაში – ანბანის რიგის მიხედვით, თუმცა სიხშირული მაჩვენებელი ამ შემთხვევაშიც ექნება სიტყვას თანდართული.



სორტირება ანბანური პრინციპით

თუ სიხშირული მაჩვენებლის[136] მიხედვით დალაგებულ პირველ გვერდს გადავხედავთ (279 შემთხვევა 1102 შემთხვევიდან), საინტერესო დასკვნების გამოტანას შევძლებთ სტატისტიკური შედეგის მიხედვით. მაგალითად, მძიმისა (536 582 შემთხვევა) და წერტილის გამოყენების სიხშირული მაჩვენებელი (247 587 შემთხვევა) იმაზე მიუთითებს, რომ მძიმე ორჯერ მეტადაა გამოყენებული, ვიდრე წერტილი, ანუ შერწყმული და ქვეწყობილი წინადადებები საკმაო სიხშირით გვხვდება ძველი ქართულის ტექსტებში. წერტილის (247 587 შემთხვევა), კითხვითი ნიშნისა (17 045 შემთხვევა) და ძახილის ნიშნის (5386 შემთხვევა) სტატისტიკა კი იმაზე მიუთითებს, რომ წართქმითი წინადადებების რაოდენობა დაახლოებით 14-ჯერ აღემატება კითხვით წინადადებას და 46-ჯერ – ძახილის წინადადებას.

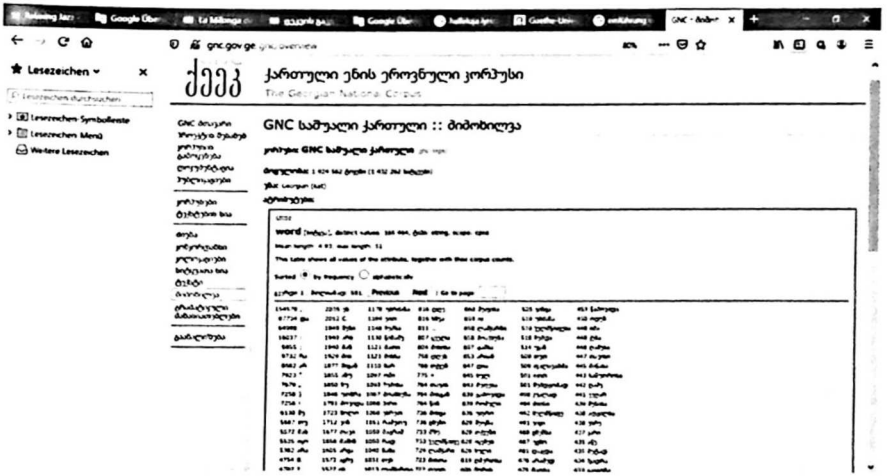
საინტერესო სურათს გვაძლევს მეტყველების ნაწილების თვალსაზრისით შედარებაც: ყველაზე მაღალი სიხშირე, როგორც მოსალოდნელი იყო, და კაუშირს გააჩნია (477 028 შემთხვევა), შემდეგ მოდის ნაცვალსახელი იგი (78 928 შემთხვევა) და კავშირი რამეთუ (60 723 შემთხვევა). საინტერესოა, რომ სიხშირული თვალსაზრისით ყველაზე მოხმარებადი ზმნა არის „*ყოფნა*“ ზმნის სხვადასხვა მწკრივის ფორმები (*არს, იყო, არიან, იყოს, იყვნეს, ხარ, იყვნენ, ვართ*), მეორე ადგილსია „*თქმა*“ (*ჰრქუა. იტყვის, იტყოდა, მიუგო, ეტყოდა, იტყოდეს, ჰრქუეს*), რომელსაც მოსდევს ზმნები *იხილა, სცა (მოვიდა, მოვიდეს, წარვიდა) ესმა, ქონა (აქუს, აქუნდა), მიცემა (მისცა), ქმნა (ყო, იქმნეს), ღირს, ბრძანება (უბრძანა), დაწყება (იწყო).* არსებითი სახელებიდან ყველაზე ხშირად ხმარებადი სიტყვებია: *ღმერთი (ღმრთისა, ღმერთმან, ღმერთსა, ღმრთისაჲ, ღმერთო), უფალი (უფალმან, უფალო, უფლისა, უფალსა, უფლისაჲ), სიტყუაჲ, კაცი (კაცთა), იესუ (იესუის), მამა (მამისა, მამათა, მამაო, მამამან, მამასა), სული (სულისა, სულსა, სულითა), ქუეყანა (ქუეყანაჲ), დღე (დღესა, დღეთა), ჯამი (ჯამსა), ქრისტე (ქრისტესითა), ძე (ძეთა, ძენი, ძისა), სახელი (სახელითა), ადგილი (ადგილსა), მეფე (მეფისა, მეფემან), ბერი (ბერმან,) ერი (ერსა, ერისა), ჴმაჲ, ძალი, ქალაქი (ქალაქსა), სიტყუაჲ (სიტყუათა), დღე (დღეს), ცხორებაჲ, იოვანე, ჯამი, ძმა (ძმათა), პირი, ჴორცი (ჴორცთა), სოფელი (სოფელსა), საქმე (საქმენი).*

[136] სტატისტიკური მაჩვენებლის მოპოვების თარიღი: 15.09.2020

საშუალი ქართულის შემთხვევაში იმავე ოპერაციას თუ გავიმეორებთ, ასეთი სურათი გვექნება: კორპუსის მოცულობა შეადგენს ტოკენებში 1 924 562 და სიტყვებში 1 432 262 ერთეულს.



სორტირების ერთეულად თუ ავირჩევთ **სიტყვას** (word) და სორტირებული სიტყვების დალაგების პრინციპად მივუთითებთ სიხშირეს, მაშინ ასეთ სურათს მივიღებთ:



პუნქტუაციის ნიშნების შეფარდება ასეთ სურათს იძლევა: ყველაზე მაღალი სიხშირით გამოირჩევა მძიმე (154 579 შემთხვევა), შემდეგ არის წერტილი (64 998

შემთხვევა). ორწერტილი თითქმის ორჯერ მეტად არის გამოყენებული, ვიდრე წერტილ-მძიმე (16037 *vs.* 9855), შემდეგ მოდის ძახილის ნიშანი (7256 შემთხვევა) და კითხვის ნიშანი (4707 შემთხვევა). ორწერტილისა და ტირეს (–) გამოყენების სიხშირე (16 037 *vs.* 3500) მიგვითითებს საშუალი ქართულის ტექსტებში დიალოგური სტრუქტურების გამოყენების ზრდაზე.

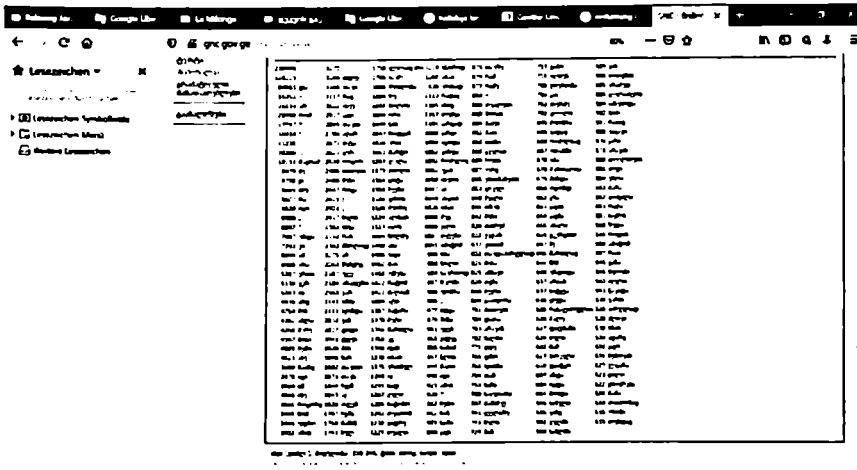
სიტყვათა მეტყველების ნაწილების მიხედვით დაჯგუფება კი ასეთ სტატისტიკურ სურათს მოგვცემს: ყველაზე მაღალი სიხშირე კვლავ და კავშირს გააჩნია (87 734 შემთხვევა), შემდეგ მოდის კითხვითი ნაცვალსახელი რა (9732) უარყოფის ნაწილაკი არ (8682), მას მოსდევს ნაცვალსახელები. საგულისხმოა, რომ პირველ, მეორე და მესამე პირებს შორის შედარებით მაღალი სიხშირით გამოირჩევა მესამე პირის ნაცვალსახელი (*მას* 5572, *იგი* 3595, *ეს* 2035, *მან* 1940, *ის* 1527), მას ოდნავ ჩამორჩება პირველი და მეორე პირის ნაცვალსახელები (*მე* 6130 და *შენ* 3232).

არსებითი სახელებიდან ყველაზე მაღალი სიხშირე აქვს სიტყვას კაცი (*კაცი* 3021, *კაცთა* 701, *კაცსა* 696, *კაცნი* 557), საქმე (*საქმე* 1496), გული (*გული* 1406, *გულსა* 955), მეფე (*მეფე* 1364, *მეფემან* 708, *მეფისა* 660, *მეფესა* 643), დღე (*დღესა* 1184, *დღე* 816, *დღეს* 798), ომი (1097), პირი (*პირი* 1066, *პირსა* 674), ლაშქარი (*ლაშქართა* 1013, *ლაშქარი* 729, *ლაშქარნი* 658), შვილი (1010), ფალავანი (*ფალავანი* 967, *ფალავანმა* 509), ღმერთი (*ღმერთმან* 885, *ღმერთსა* 619, *ღმერთი* 532), წიგნი (845), ბე (821), ცხენი (*ცხენი* 736, *ცხენსა* 488), მზე (*მზე* 733, *მზისა* 573), ჭელმწიფე (*ჭელმწიფე* 733, *ჭელმწიფესა* 519, *ხელმწიფე* 492, *ჭელმწიფემან* 471), ამბავი (720), როსტომი (*როსტომ* 709), ქალი (688), ტახტი (*ტახტი* 679, *ტახტსა* 430), ჟამი (*ჟამსა* 657), სული (626), ლხინი (589), ხელი (ხელი 573) *vs.* ჭელი (*ჭელი* 535, *ჭელთა* 465), ჯარი (526), მადლი (474), სისხლი (474), დევი (472), ლახტი (470), გზა (467), ფრიდონი (*ფრიდონ* 465), სახელი (462), მტერი (*მტერთა* 457), ქება (448), ღამე (*ღამესა* 448, *ღამე* 442), მიწა (*მიწასა* 445), საწუთრო (*საწუთროსა* 443), ადგილი (*ადგილსა* 438), კარი (437), ქალაქი (*ქალაქსა* 433, *ქალაქი* 425), საუბარი (430), წამი (*წამსა* 430), დიდებული (*დიდებულნი* 427), სახლი (*სახლსა* 422), მანუჩარი (*მანუჩარ* 419), ყმა (*ყმა* 409).

ზმნებიდან ყველაზე მაღალი სიხშირით გამოირჩევიან ყოფნა (იყო 5535, არის 2535, არს 1940, ხარ 1110, იყოს 690, არიან 653, იყენეს 628), თქმა (უთხრა 1846, თქვა 1248,), სვლა (მოვიდა 1791, მივიდა 976, წავიდა, გამოვიდა 639, მოვიდეს 583, წამოვიდა 453), ბრძანება (უბრძანა 1178, ბრძანა 527, უბძანა 519), მოხსენება (მოახსენა 1067, მოაჯსენა 658), ნახვა (ნახა 1040, ნახეს 466), მიცემა (მისცა 736), ქონა (ჰქონდა 715), ქმნა (შეიქნა 629, იქმნა 424), სმენა (ესმა 550), ცოდნა (ეცი 491), დაჯდომა (დაჯდა

481), ადგომა (ადგა 467), კადრება (ჰკადრა 434), გაგება (გაიგონა 433), გათენება (გათენდა 417).

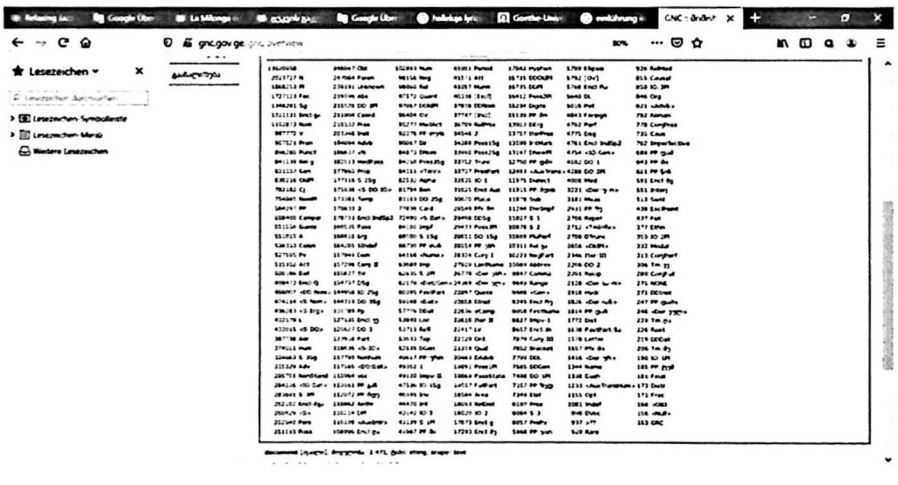
ახალი ქართულის არჩევის შემთხვევაში სურათი ასეთი იქნება: ტოკენსა (2 333 477) და სიტყვებს (2 108 370) შორის სხვაობა საგრძნობლად არის შემცირებული. მძიმის გამოყენების სიხშირე კლავ თითქმის ორჯერ აღემატება წერტილის გამოყენების სიხშირეს (230 995 *vs.* 128 221), კითხვის ნიშნის გამოყენების სიხშირე ოდნავ აღემატება ძახილის ნიშნის გამოყენების სიხშირეს (17 747 *vs.* 16 414), რაც დიალოგების გამოყენების სიხშირეზე მიუთითებს. საგულისხმოა, რომ და კავშირი აქაც ყველაზე მაღალი სიხშირით გამოირჩევა, თუმცა პროცენტული თვალსაზრისით ბევრად ჩამოუვარდება ძველ და ახალ ქართულს. აქვე უნდა აღვნიშნოთ, რომ კავშირებს შორის პირველ ადგილზე დგას რომ (20 969) და მაგრამ (10 111), რაც რთული ქვეწყობილი წინადადებების სინტაქსური გაფორმების განვითარებაზე მიუთითებს. ნაწილაკებიდან კლავ გამოირჩევა უარყოფის ნაწილაკი არ (26519). პირველ და მეორე პირის ნაცვალსახელებს შორის ისევ ლიდერობს პირველი პირის ნაცვალსახელი (*მე* 9979 *vs. შენ* 4754). ფუნქციური ელემენტებიდან თვალშისაცემია მოდალური სიტყვის უნდა სიხშირე (7987). რაც შეეხება არსებით სახელებს, კლავ ლიდერობს კაცი (2917), ხოლო ზმნებიდან იყო (9520).



კორპუსი საშუალებას იძლევა სორტირება მოვახდინოთ გრამატიკული მახასიათებლების თვალსაზრისითაც. დავიწყოთ ძველი ქართულის კორპუსის ანალიზით. როგორც ქვემოთ მოყვანილი ცხრილიდან ჩანს, ყველაზე მაღალი სიხშირით გამოირჩევა სახელი (2022727 შემთხვევა), შემდეგ მოდის ზმნა (987772), რაც ბუნებრივიცაა

– ზმნას, როგორც წესი, რამდენიმე სახელი შეეწყობა. რაც შეეხება გვარს, მოქმედებითი და ვნებითი გვარის ფორმების შეფარდება ასეთია: მოქმედებითი 515102 *vs.* ვნებითი 169535. საგულისხმოა, რომ მრავლობითის ფორმების გამოყენება ბევრად აღემატება მხოლობითი რიცხვის გამოყენების შემთხვევებს (1868253 *vs.* 1348281). ბრუნვის თვალსაზრისით სახელობითი ბრუნვის ფორმები ჭარბობს (1192873), მეორე ადგილზეა ნათესაობითი ბრუნვა (831337). ასევე საგულისხმოა, რომ სახელობითში მდგარი სუბიექტისა და მოთხრობითში მდგარი სუბიექტის სიხშირე დაახლოებით თანაბარია: S:NOM = 474 114 *vs.* S:ERG =36 283. მიცემითში მდგარი სუბიექტის სიხშირე ბევრად დაბალია: S:DAT =72 490. ცალკე აღებული სახელობითის, მოთხრობითისა და მიცემითის ბრუნვის ფორმების შეფარდება ასეთია: სახ.: 1192873, მოთხ.: 168816, მიც.: 500186. თანდებულიანი ფორმების სიხშირული მონაცემია 131789. მათგან ყველაზე გავრცელებულ -გან და -მდე თანდებულიანი ფორმებია (113161 *vs.* 112072). შემდეგია -თვის და -თან თანდებულიანი ფორმები (95276 ცს. 66730). შედარებით დაბალი სიხშირით დასტურდებიან თანდებულები: -დამი (247), ქვეშ (185), -მდე (64), -კენ (3).

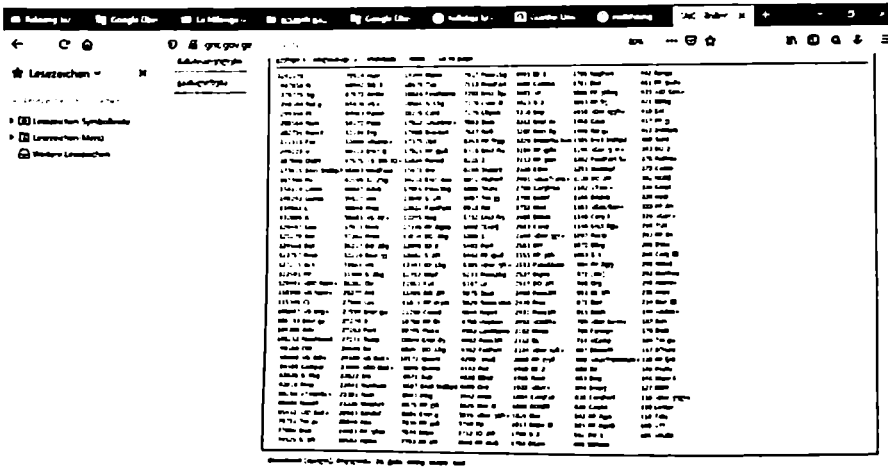
რაც შეეხება მწკრივებს, ნამყოს ფორმების სიხშირე მაღალია 387738, მეორე ადგილზეა აწმყოს მწკრივი – 210332.



საშუალი ქართულის მონაცემები ასეთია: აქაც, ყველაზე მაღალი სიხშირით სახელი გამოირჩევა: სახელი (497856 შემთხვევა) *vs.* ზმნა (249220). რაც შეეხება გვარს, მოქმედებითი და ვნებითი გვარის ფორმების შეფარდება ასეთია: მოქმედებითი 123 215 *vs.* ვნებითი 54 172. ძველი ქართულისაგან განსხვავებით მრავლობითის ფორმების გამოყენება ბევრად ჩამორჩება მხოლობითი რიცხვის ფორმებს (მრ. 294304 *vs.* მხ.

376775). ბრუნვის თვალსაზრისით აქაც სახელობითი ბრუნვის ფორმები ჭარბობს (290564), მეორე ადგილზეა ნათესაობითი ბრუნვა (129597). ასევე საგულისხმოა, რომ სახელობითში მდგარი სუბიექტისა და მოთხრობითში მდგარი სუბიექტის სიხშირე დაახლოებით თანაბარია: S:NOM 118 598 *vs.* S:ERG=106 007. მიცემითში მდგარი სუბიექტის სიხშირე აქაც ბევრად დაბალია: S:DAT 24=109. ცალკე აღებული სახელობითის, მოთხრობითისა და მიცემითის ბრუნვის ფორმების შეფარდება ასეთია: სახ.: 290564, მოთხ.: 53104, მიც.: 124550. თანდებულიანი ფორმების სიხშირული მონაცემია 122591. მათგან ყველაზე გავრცელებულია -ურთ (19683), -დან თანდებულიანი ფორმები (17025). შემდეგ მოდის -მდის (13196) და -თვის თანდებულიანი ფორმები (11871). შედარებით დაბალი სიხშირით დასტურდებიან თანდებულები -წინ (156), -ებრ (93), -და (18).

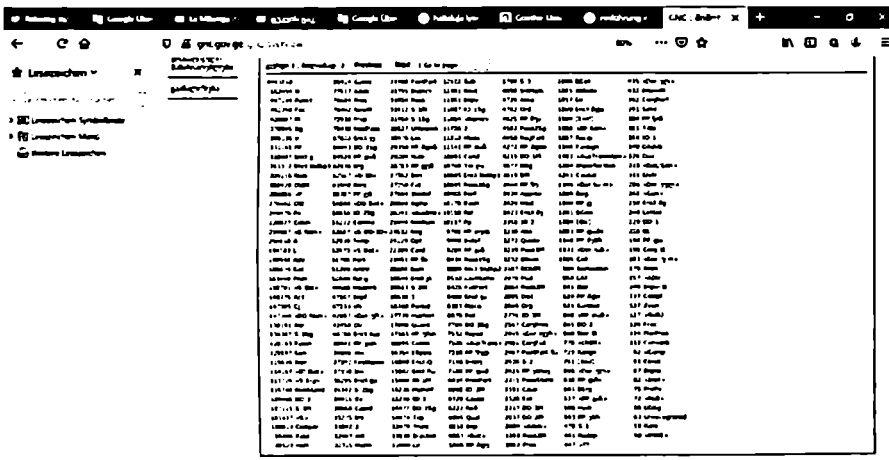
რაც შეეხება მწკრივებს, ნამყოს ფორმების სიხშირე მაღალია 125279, მეორე ადგილზეა აწმყოს მწკრივი – 28840.



ახალი ქართულის მონაცემები იმავე პარამეტრების მიხედვით ასეთია: აქაც, ყველაზე მაღალი სიხშირით გამოირჩევა სახელი (552454 შემთხვევა) ეიდრე ზმნა (369136). რაც შეეხება გვარს, მოქმედებითი და ენებითი გვარის ფორმების შეფარდება ასეთია: მოქმედებითი 148375 *vs.* ენებითი 95946. აქაც, ძველი ქართულის მსგავსად მრავლობითის ფორმების გამოყენება ჭარბობს მხოლობითი რიცხვის ფორმებს (მრ. 420867 *vs.* მხ. 378095). ბრუნვის თვალსაზრისით აქაც სახელობითი ბრუნვის ფორმები ჭარბობს (309218), მეორე ადგილზეა მიცემითი ბრუნვა (186074), მესამეზე – ნათესაო-

ბითი 120597. მოთხრობითი ფორმების სიხშირე საგრძნობლად დაბალია (62936). ასევე საგულისხმოა, რომ სახელობითში მდგარი სუბიექტის სიხშირე თითქმის ორჯერ მეტია, ვიდრე მოთხრობითში მდგარი სუბიექტის სიხშირე: S:NOM=204087 *vs.* S:ERG= 111726. მიცემითში მდგარი სუბიექტის სიხშირე აქაც ბევრად დაბალია: S:DAT= 52673. ცალკე აღებული სახელობითის, მოთხრობითისა და მიცემითი ბრუნვის ფორმების შეფარდება ასეთია: სახ.: 309218, მოთხ.: 62936, მიც.: 186074. თანდებულიანი ფორმების სიხშირული მონაცემია 331165. მათგან ყველაზე გავრცელებულ -დან (64528) და -კენ თანდებულიანი ფორმები (59387). შემდეგ მოდის -ეით (39441) და -ქვეშ თანდებულიანი ფორმები (28783). შედარებით დაბალი სიხშირით დასტურდება -ებრ (503), -წინ (384), -და (199).

რაც შეეხება მწკრივებს, ნამყოს ფორმების სიხშირე აქაც მაღალია 138 191, მეორე ადგილზეა აწმყო – 76689, მესამეზე კი მომავლის მწკრივი – 27250.



აქ განხილული სტატისტიკური მონაცემები **დისტანციური კითხვის**[137] საშუალებას იძლევა: ძველი ქართულის ლექსიკა მიგვანიშნებს რელიგიური ხასიათის ტექსტებზე, საშუალი ქართულის ლექსიკისაგან განსხვავებით, რომელიც საერო ხასიათის ტექსტებზე მიუთითებს. ფუნქციურ ელემენტთა სიხშირე, როგორიცაა მაგალითად კავშირები, წინადადებების სინტაქსური დახასიათების საშუალებას იძლევა.

[137]**დისტანციური კითხვის**, როგორც ფენომენის შესახებ დაწვრილებით იხ.: რ. ხალვაში, დიგიტალური ჰუმანიტარიის შესავალი, ბათუმი, 2018, გვ. 131-140.

სინტაქსური სტრუქტურების დიაქრონიულ ჭრილში ანალიზი კი წინადადების სტრუქტურების გართულების ან გამარტივების მაჩვენებელია.

აქვე საგანგებოდ გვინდა შევეხოთ სტატისტიკური მონაცემების პროცენტულობასთან მიმართების საკითხს. როდესაც ამა თუ იმ სიტყვაფორმის სტატისტიკური მაჩვენებლები მოგვყავს, საქმე გვაქვს აბსოლუტურ სტატისტიკასთან, ანუ მოცემულ კორპუსში მოცემული ელემენტის პოვნიერების აბსოლუტურ სიხშირესთან. მაგრამ თუ ჩვენ კორპუსში (ან კორპუსის რომელიმე ტექსტში) ამა თუ იმ ფორმის არა პოვნიერების, არამედ მატება-კლების პროცენტულობას ვიკვლევთ (მაგალითად, განსაზღვრელობითი არტიკლის ან ნაცვალსახელის გამოყენების სიხშირეს დიაქრონიული თვალსაზრისით ძველ, საშუალ და ახალ ქართულში), მაშინ დაუშვებელია აბსოლუტური სტატისტიკური მონაცემების შედარება. საჭიროა ე.წ. შეფარდებითი სტატისტიკის გამოთვლა, ანუ, სტატისტიკური მონაცემის პოვნიერების სიხშირის გამოთვლა შესაბამისი სუბკორპუსების მოცულობასთან მიმართებაში. ამ ფენომენს **დაბალანსების ფაქტორი** ეწოდება და ქართული ენის დიაქრონიული კვლევის თვალსაზრისით არსებითი და გადამწყვეტია. მაგალითად, საშუალი ქართულის რესურსების მოცულობა გაცილებით ნაკლებია ძველი და ახალი ქართულის მოცულობასთან შედარებით. ძველი ქართული ენის კორპუსი კი თავის მხრივ დაახლოებით 4,2-ჯერ უფრო დიდი, ვიდრე საშუალი ქართულის კორპუსი. ხოლო თანამედროვე ქართული ენის კორპუსი დაახლოებით 2,9-ჯერ უფრო პატარაა, ძველი ქართულის კორპუსთან შედარებით.[138]

საგანგებოდ გვინდა აღვნიშნოთ, რომ აქ მოყვანილი სტატისტიკური მონაცემები არ არის უცცელელი. ვინაიდან ქართული ენის ეროვნული კორპუსი განვითარებად რესურსს წარმოადგენს, მისი კონტენტი და შესაბამისად მოცულობაც მუდმივად იზრდება. აქედან გამომდინარე, ზემოთ მოხმობილი სტატისტიკა მეთოდურ ოპერაციას წარმოადგენს და არ ისახავს მიზნად აღნიშნული რესურსის საბოლოო სტატისტიკურ შეფასებას.

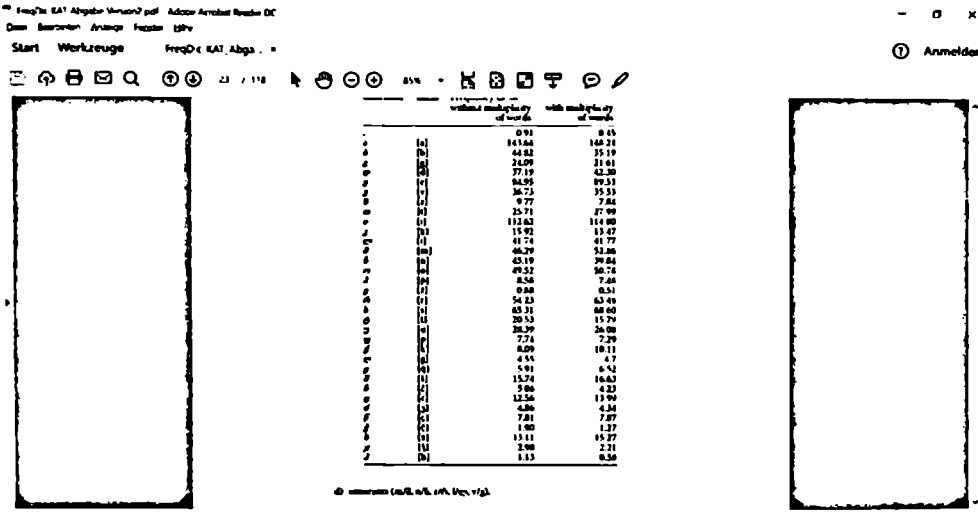
ზემოთ უკვე აღვნიშნეთ, რომ სიტყვების, როგორც X ცვლადის Y ფუნქციის გასომვა შესაძლებელია მთელი ტექსტის, ან ტექსტის გარკვეული მონაკვეთის ფარგლებში. ტექსტის სტატისტიკური დამუშავების ეს შესაძლებლობა, როგორც წესი, ინტეგრირებულია თანამედროვე კორპუსების სამომხმარებლო ინტერფეისში.

---

[138] ეს საკითხი მეთოდური თვალსაზრისით საგანგებოდ არის განხილული ნაშრომში: Mariam Kamarauli, The Nominal Domain in Georgian (A diacronic analysis), Diss. Frankfurt, 2018 (in press).

მაგალითად. ლაიფციგის უნივერსიტეტის მიერ შექმნილ ქართული ენის კორპუსში ინტეგრირებულია ტექსტის სტატისტიკური კვლევის ოპერაციული სისტემა არა მხოლოდ გრაფემების, სიტყვაფორმების, ან მეტყველების ჯგუფების სიხშირის კვლევის შესაძლებლობით, არამედ სიტყვათა ფორმალური მოცულობის (გრაფიკური სიგრძის) კვლევის შესაძლებლობითაც კი.

როგორც ქვემოთ მოყვანილი სტატისტიკური სურათიდან ჩანს, ყველაზე მაღალი სიხშირით, რასაკვირველია, ხმოვნები ხასიათდებიან ასეთი თანმიმდევრობით: ა, ი, ე, ო, უ; ყველაზე დაბალი სიხშირით კი – თანხმოვანი ჟ.



რაც შეეხება სიტყვათა სიგრძის სტატისტიკას, ძალიან საინტერესო სურათი გვაქვს ამ თვალსაზრისით, რაც რამდენიმე ლინგვისტური ხასიათის, მათ შორის სიტყვის დეფინიციის საკითხის გადაწყვეტას უკავშირდება. ლაიფციგის ქართული ენის კორპუსი ე.წ. ვებკორპუსი გახლავთ – რესურსები, რომელიც ამ კორპუსში იქნა ინტეგრირებული, ონლაინრესურსებს წარმოადგენს და საერთაშორისო გლობალური ქსელიდან არის აღებული. ინტერნეტგვერდებიდან გამოყენებულია შემდეგი გვერდები: ka.wikipedia.org, news.ge, www.opentext.org.ge, www.ambebi.ge, www.interpressnews.ge, topnews.mediamall.ge, ici.ge, halamdrid.ge, life.mediamall.ge, geonews.ge და rss.ge. გარდა ამ გვერდებისა, რესურსები აღებულია სოციალური ქსელებიდანაც.

როგორც ქვემოთ მოცემული სტატისტიკური კვლევის სურათი გვიჩვენებს, აღნიშნული კორპუსის სტატისტიკურმა ანალიზმა ენის ეკოლოგიის თვალსაზრისით საინტერესო ფენომენებთან ერთად (როგორიცაა ნორმების დარღვევის შემთხვევები,

გრაფიკული სახესხეაობები უცხო ენიდან შემოტანილი სიტყვების დაწერილობის შემთხვევაში და ა.შ.) სხვა სახის პრობლემებიც წარმოაჩინა, რომელიც უპირატესად „ზეპირმეტყველების წერილობით რეალიზების“ ფენომენთან არის დაკავშირებული. აქ, უპირველეს ყოველისა, მხედველობაში გვაქვს ემოციურობის ხარისხის გადმომცემი თანამედროვე – გრაფიკული საშუალებები:

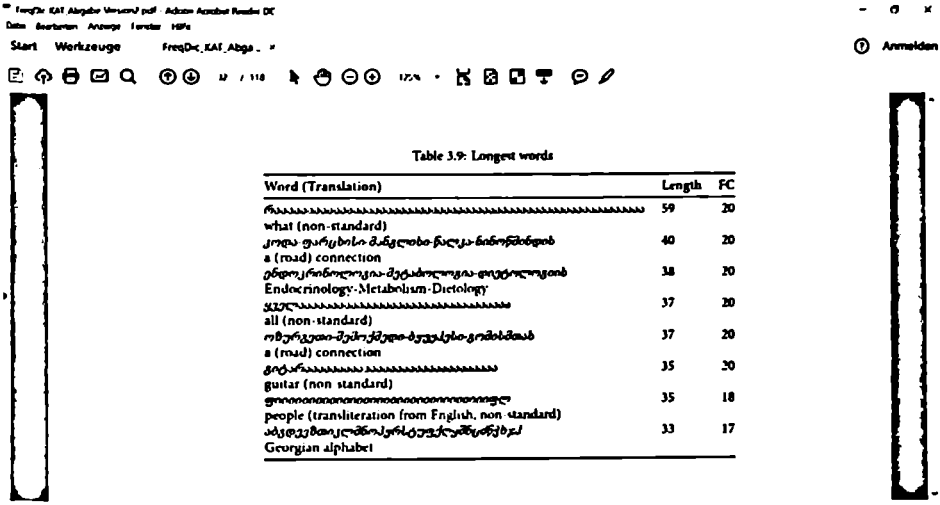


Table 3.9: Longest words

Word (Translation)	Length	FC
რ	59	20
what (non-standard)		
კოდა-ფარცხისა-მანქანახა-ნაღა-ბინორშინდან	40	20
a (road) connection		
ენდოკრინოლოგია-მეტაბოლიზმი-დიეტოლოგიის	38	20
Endocrinology-Metabolism-Dietology		
ყველ	37	20
all (non-standard)		
ოზურგეთი-შეხვედრა-ბუღალტრი-გომხმას	37	20
a (road) connection		
გიტარა	35	20
guitar (non standard)		
ფოიიი

7. ქართული ენის ეროვნული კორპუსის სახელმწიფოებრივი მნიშვნელობა

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

1) 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;

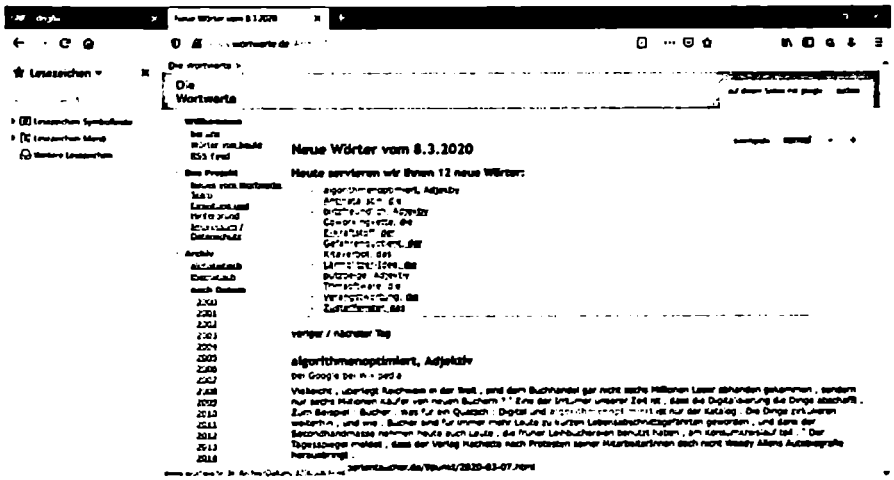
2) ქართული ენის **სრულფასოვანი ფუნქციონირების უზრუნველყოფა.**

   აქ იგულისხმება როგორც ტექნიკური (სპელჩეკერისა და პარსერის დახვეწა ქართული ენისათვის), ისე იურიდიული ასპექტი, როგორიცაა, მაგ. ქართული ენის, როგორც სახელმწიფო ენის დიდ მონაცემთა ბაზაზე (სტატისტიკური ლინგვისტიკის მეთოდების გამოყენებით) აპრობირებული სასწავლო დიგიტალური პროგრამების შექმნა, მათ შორის არაქართველი მოსახლეობისთვისაც.

აქვე გესურს, ხაზი გაეუსვათ 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.

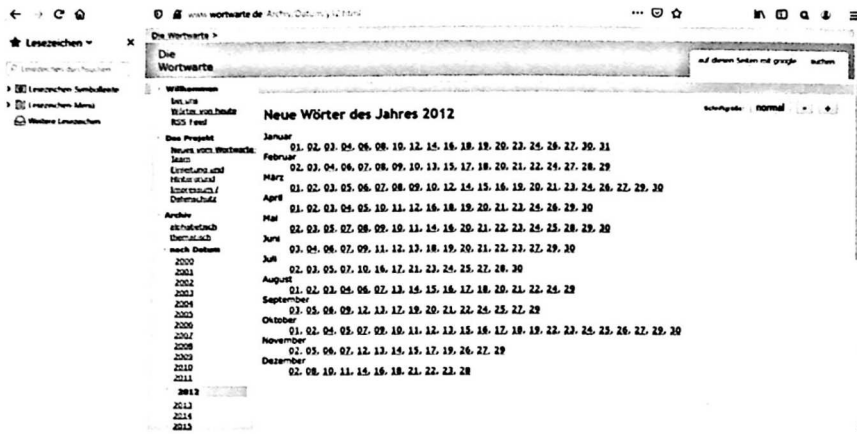
ენის ეროვნული კორპუსის სახელმწიფოებრივი მნიშვნელობის საილუსტრაციოდ მოვიყვანთ ერთ მაგალითს გერმანული გამოცდილებიდან: ეს გახლავთ უახლესი ნასესხობებისა და ახალი კომპოზიტების ნორმალიზაციისა და სტანდარტიზაციის საკითხი. უკვე ათწლეულზე მეტია (2000 წლიდან), რაც მანჰაიმის გერმანული ენის ინსტიტუტში IDS (Institut für deutsche Sprache) ფუნქციონირებს გერმანული ენის "სიტყვის სადარაჯო" Die Wortwarte (www.wortwarte.de), რომლის იდეის ავტორი და განმახორციელებელი გახლავთ ლოთარ ლემნიცერი. ამ სამსახურის მიერ ყოველდღიურად მიმდინარეობს გერმანული გაზეთების ელექტრონული ვერსიების „შემოწმება" – მთელი იმ დღის მასალა პროგრამული უზრუნველყოფის შედეგად მიემარ-

თება გერმანულ 4 მილიარდსიტყვიან კორპუსს, ხდება ავტომატური ამოფილტრა იმ დღეს პირველად გამოყენებული და „კორპუსისათვის უცნობი“ სიტყვებისა, რომლებიც ნორმალიზაციისათვის გადაეგზავნება ლინგვისტთა სპეციალურ ჯგუფს; აღნიშნული ჯგუფი განიხილავს სიტყვის დაწერილობის სისწორეს და ადგენს/განმარტავს მის მნიშვნელობას. სიტყვის მოხმარების მიზანშეწონილების შემთხვევაში, ხდება მისი სტანდარტიზაცია – ახალი სიტყვა თავის ადგილს პოულობს კორპუსში – შესაბამისი კონტექსტი ჩაერთვის კორპუსში, ხოლო ნორმატიული ფორმა შედის ლექსიკონში. ასე, ყოველდღიურად იზრდება გერმანული ეროვნული კორპუსი ლინგვისტთა მეთვალყურეობის ქვეშ. ქვემოთ მოცემულ სურათზე ასახულია, მაგალითად, 2020 წლის 8 მარტის ახალი სიტყვები.

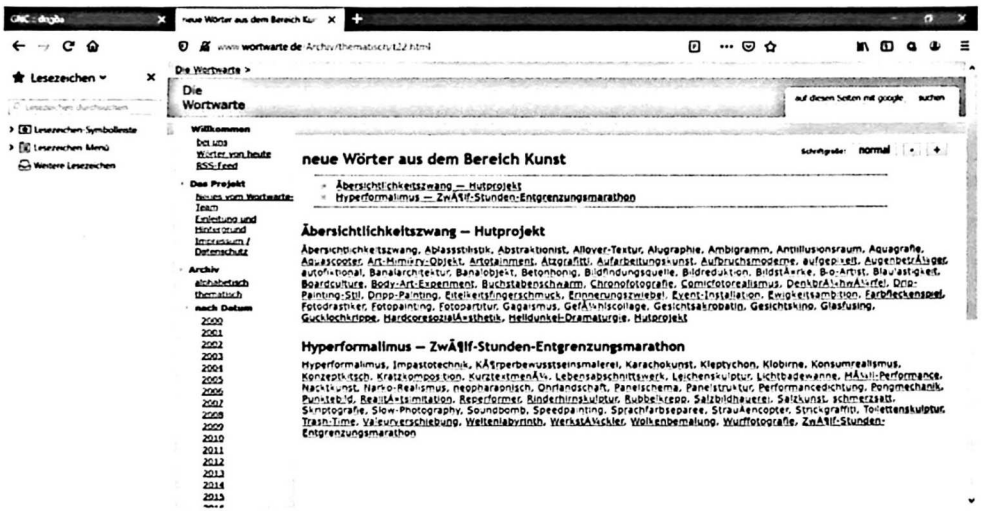


ენისადმი სახელმწიფოებრივი დამოკიდებულების შესანიშნავი ნიმუშია თავად ლოთარ ლემნიცერის სიტყვები: „არსებობს ამინდის ბიურო (Wetterwarte), რომელიც აკვირდება ჰაერის მასების მოძრაობას და პროგნოზირებს ამინდს. არსებობს ობსერვატორია (Sternwarte), რომელიც აკვირდება ციური სხეულების შექმნას, მოძრაობასა და გაქრობას. ჩვენ შევქმენით „სიტყვის სადარაჯო“ (Wortwarte, აქ სიტყვათა თამაშია. შდრ. Wetterwarte და Sternwarte), რათა დავაკვირდეთ სიტყვების შექმნას, გერმანული ლექსიკის განვითარებას და ვამცნოთ საზოგადოებას მათი არსებობის შესახებ. ჩვენ დამკვირვებლები ვართ, ენის დამკვირვებლები.... ჩვენი დაკვირვების ინტერპრეტაცია კი მათთვის მიგვინდია, ვისაც ეს ეკალება“.

2000 წლიდან მოყოლებული თქვენ შეგიძლიათ თვალი მიადევნოთ ახალი სიტყვების დაბადებას, ნორმალიზაციას და სტანდარტიზაციას.



მომხმარებელს შეუძლია ასევე თემატურად მოძებნოს სიტყვები, როგორც, მაგალითად, ქვემოთ არის მოცემული – ახალი სიტყვები ხელოვნების სფეროდან:



გერმანიაში ყოველწლიურად ცხადდება კონკურსი „წლის სიტყვის“ გამოსაცნობად. ეს კონკურსი პირველად 1971 ჩაატარა გერმანული ენის სახოგადოებამ (Gesellschaft für deutsche Sprache)[140] და 1977 წლიდან ტრადიციად იქცა. 1991 წლიდან ნომინაციების სიას შეემატა წლის „ანტისიტყვა“ (გერმ.: Unword), 2008 წლიდან კი – წლის ახალგაზრდული სიტყვა.

[140] https://gfds.de/

საგულისხმოა, რომ წლის სიტყვები საზოგადოების ინტერესისა და განწყობის ამსახველია. მაგალითად, 2001 წლის სიტყვა იყო „11 სექტემბერი“, 2005 წლის – „(გერმანიის ფედერაციული რესპუბლიკის) კანცლერი“, 2007 წლის – „კლიმატური კატასტროფა“, 2008 წლის – „ფინანსური კრიზისი“, 2014 წლის – „დიდი კოალიცია“ (GroKo), 2015 წლის – „ლტოლვილები“, 2019 წლის სიტყვად კი გამოცხადდა Respektrente „პატივისცემის (ნიშნად დანიშნული) პენსია“. ეს სიტყვა პირველად 2019 წელს გამოჩნდა გერმანულ ენაში და მისი გამოჩენა უკავშირდება გერმანიის ფედერაციული რესპუბლიკის შრომის მინისტრის ჰუბერტუს ჰაილის (Hubertus Heil) სახელს, რომელმაც შემოიტანა წინადადება, რომ მათთვის, ვისაც შრომით ბიოგრაფიაში არ დაუგროვდება შრომის კანონმდებლობით გათვალისწინებული საპენსიო წლების რაოდენობა, მათ მაინც დაენიშნოთ მაღალი პენსია მათ მიერ განვლილი ცხოვრების (სოციალური აქტივობის) დაფასების მიზნით. 2019 წლის ანტისიტყვად კი გამოცხადდა Klimahysterie „კლიმატური ისტერია“. ეს სიტყვა მეცნიერებმა დისკრიმინაციულ სიტყვად მიიჩნიეს იმ ადამიანებთან მიმართებაში, რომლებიც აქტიურად არიან ჩართული კლიმატის ცვლილებებით გამოწვეული კატასტროფების გამოვლენასა და მის წინააღმდეგ აქტიურ ბრძოლაში.

სიტყვის ანტისიტყვად გამოვლენის მიზეზი ხშირად მოცემული სიტყვის არამისხობრივი გამოყენება ან პოლიტიკური არაკორექტულობაა. მაგალითად, 2010 წლის ანტისიტყვად დასახელდა ზედსართავი სახელი alternativlos[141] „უალტერნატივო“, ვინაიდან ეს სიტყვა საზოგადოებას პესიმისტურად განაწყობს და თავიდანვე გამორიცხავს დებატების მიზანშეწონილებას.

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, რაც კიდევ ერთხელ დაადასტურა კოვიდ-19-ის პანდემიის პირობებში სავალდებულო დისტანციური სწავლების აუცილებლობამ.

თავისი მნიშვნელობიდან გამომდინარე საგანგებოდ გვინდა შევეხოთ ქართული ენის ეროვნული კორპუსის მეცნიერული გამოყენების ასპექტს, განსაკუთრებით კი ეროვნული კორპუსის მნიშვნელობას დიგიტალური ლექსიკოლოგიის განვითარების თვალსაზრისით.

---

[141] ინგლისურ ენაში მის შესატყვისად გამოიყენება აკრონიმი „TINA“ (*there is no alternative*).

8. ქართული ენის ეროვნული კორპუსის მეცნიერული გამოყენების სფეროები

ენის ეროვნული კორპუსის არსებობა ერთ-ერთი წინაპირობაა საქართველოში ემპირიული ცნობიერების განვითარებისათვის სამეცნიერო სფეროში. კორპუსების არსებობა განსაკუთრებით მნიშვნელოვანია კვლევის თანამედროვე მეთოდების წარმატებით გამოყენების თვალსაზრისით – იგი სერიოზულ ემპირიულ საფუძველს უქმნის ქართველოლოგიურ კვლევებს და ხელს უწყობს კორპუსზე ორიენტირებული კვლევების დანერგვას ფილოლოგიაში. რაც შეეხება პრაქტიკულ შედეგებს, ენის ეროვნული კორპუსი საშუალებას გვაძლევს ქართული ენის სტრუქტურა, ისტორია და განვითარების ტენდენციები თანამედროვე ტექნოლოგიური საშუალებების გამოყენებით ვიკვლიოთ. გარდა ამისა, ეროვნული კორპუსის არსებობა ერთ-ერთი ხელშემწყობი ფაქტორია იმისათვის, რომ სახელმწიფო ენისათვის უმნიშვნელოვანესი დარგი – ლექსიკოგრაფია სრულიად ახალ, თვისებრივად განსხვავებულ საფეხურზე ავიყვანოთ. აქ ჩვენ ვგულისხმობთ როგორც ქართული ენის განმარტებითი ლექსიკონის შევსებისა და დახვეწის მექანიზმის შემუშავებას თანამედროვე ინოვაციური მეთოდების გამოყენებით, ისე ორ- და მრავალენოვანი ლექსიკონების შექმნის შესაძლებლობას. 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 (TITUS-ის და ARMAZI-ის) მოხმარების გამოცდილებას თუ გავითვალისწინებთ, ენის კორპუსის მომხმარებელი საზოგადოების ძალიან მცირე ნაწილია. ესენი არიან სამეცნიერო ფილოლოგიური საზოგადოების წარმომადგენლები (და ისიც მცირე), აგრეთვე ფილოლოგიური პროფილის სტუდენტთა ძალიან შეზღუდული რაოდენობა. ეროვნული კორპუსის, უფრო სწორედ, ეროვნულ კორპუსზე დაფუძნებული პროგრამების მომხმარებელი შეიძლება და უნდა გახდეს კიდეც 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 იქნეს გათვალისწინებული. გარდა დიდი სახელმწიფოებისა, როგორებიცაა ამერიკა, გერმანია, საფრანგეთი, ეროვნული კორპუსი შექმნილი აქვთ და მის შესაქმნელად ძალისხმევა არ დაიშურეს, უპირველეს ყოვლისა, სახელმწიფოებრიობის მქონე მცირე ერებმა. დასახელებული კორპუსების გარდა, დღეისათვის არსებობს რუსული, უნგრული, ჩეხური, ხორვატული, პოლონური, ბულგარული, ალბანური, ბერძნული, თურქული, სლოვაკური, (აღმოსავლეთ) სომხური ეროვნული კორპუსები.

თავი მეხუთე

ქართული ანბანის საერთაშორისო სტანდარტიზაცია უნიკოდში

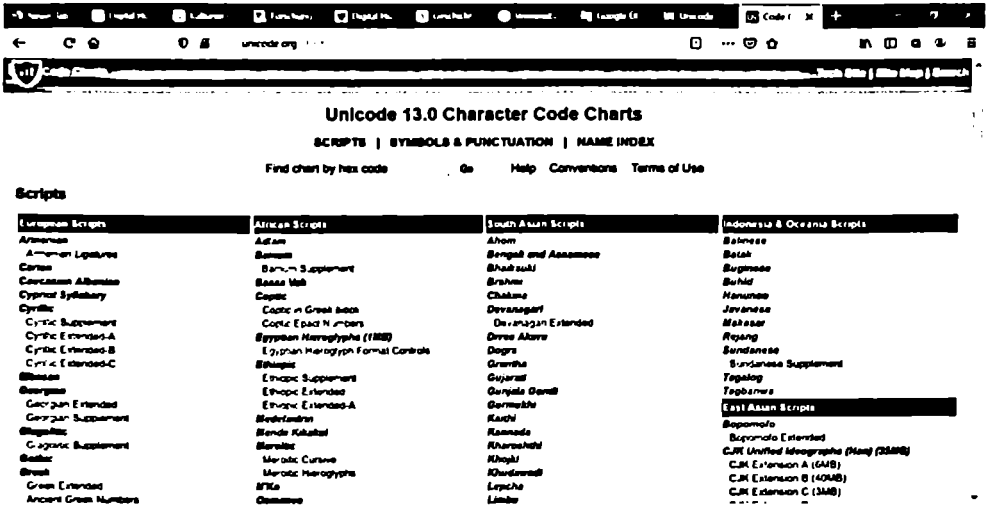
1. უნიკოდი და ქართული ანბანის საერთაშორისო კოდირება

ქართული ენის ეროვნული კორპუსი ეყრდნობა ქართული ენისათვის უკვე არსებულ რამდენიმე კორპუსის მასალას – TITUS, ARMAZI, GEKKO, GDC, Opentext, Lib.de, რომლებიც ერთმანეთისაგან განსხვავდებიან ისეთი არსებითი მახასიათებლების მიხედვით, როგორიცაა კორპუსის სტრუქტურა, ლინგვისტური ანოტაცია, მეტამონაცემები, სამომხმარებლო ინტერფეისი და ა.შ. ქართული ენის ეროვნულ კორპუსში მათი გაერთიანება აუცილებლობით მოითხოვდა, ერთი მხრივ, მასალების გაერთმნიშვნელიანებას საერთაშორისო სტანდარტების გათვალისწინებით და მეორე მხრივ, საერთო ტექნოლოგიური ჩარჩოს კონცეპტუალიზაციას.

დიგიტალურ ჰუმანიტარიაში **კოდირება (character encoding)** ნიშნავს ლინგვისტური ტექსტის გადაყვანას კომპიუტერული მეხსიერებისათვის გასაგებ ნიშნებში – ბიტებში. კოდირების პირველი, ე. წ. 8-ბიტიანი კოდირების საერთაშორისო სტანდარტი (ISO-8859) სულ მალე 16-ბიტიანმა კოდირების სტანდარტმა შეცვალა, რომელიც სტანდარტების საერთაშორისო კონსორციუმმა უნიკოდმა (Unocode Consortium) 1991 წელს გამოაქვეყნა. ეს გახლავთ მანქანურად წაკითხვადი ტექსტის კოდირებისა და დეკოდირების სტანდარტული სისტემა, რომელიც საერთაშორისო სტანდარტის ISO-10646-ის ნაწილს წარმოადგენს და კოდირების 16-ბიტიან სისტემას იყენებს.

უნიკოდი, როგორც საერთაშორისო კონსორციუმი (საიდენტიფიკაციო კოდით 501(c)(3)), არაკომერციული ორგანიზაციაა, რომელიც მიზნად ისახავს ენათა დიგიტალური გამოყენების შეუფერხებელ მიმოქცევას საერთაშორისო სივრცეში.

მსოფლიოს ხუთ კონტინენტზე 7000-მდე ენა გამოიყენება, რომელთაგანაც მხოლოდ 100 ენისათვის არსებობს საერთაშორისო გამოყენების დიგიტალური ფორმატი – კომპიუტერში, ტელეფონში, სმარტფონში, აიფონში და სხვა თანამედროვე ტექნიკურ მოწყობილობებში დანარჩენი ენებისათვის დიგიტალური მხარდაჭერა არ არსებობს.



უნიკოდის მიერ პროგრამებისა და მოწყობილობებისათვის შექმნილი ტექნოლოგიური მხარდაჭერა სხვადასხვა ენებთან მიმართებაში განსხვავებულ დონეს გეიჩევნებს და სამ ჯგუფად იყოფა:

1. ენის საბაზისო მოხმარების მხარდაჭერა (Basic language support)
2. ენის შინაარსობრივი მოხმარების მხარდაჭერა (Content language support)
3. ენის სამომხმარებლო დონეზე მხარდაჭერა (UI language support)

ენის საბაზისო მოხმარების მხარდაჭერა გამოიხატება შემდეგ პარამეტრებში: პროგრამები და ტექნიკური მოწყობილობები აღჭურვილია უნიკოდის მიერ კოდირებული საგანგებო ენობრივი ნიშნების ბადით და მათი გამოყენებისათვის აუცილებელი შესაბამისი ალგორითმებით, მომხმარებელს საშუალება ეძლევა გამოიყენოს ამ ენისათვის შექმნილი შრიფტების ნაირსახეობა, შექმნილია შრიფტების კლავიატურაზე განლაგების სისტემა. ამ სახის მხარდაჭერის შემთხვევაში ინფორმაციის გადატანა ერთი ტექნიკური მოწყობილობიდან მეორეზე, ანუ ინფორმაციის უხარვეზო გაცვლა და მათი ტექსტური დამუშავება პრობლემას არ წარმოადგენს.

უნიკოდის მიერ ენათა ტექნოლოგიური მხარდაჭერის უფრო მაღალ საფეხურს წარმოადგენს **ენის შინაარსობრივი მოხმარების მხარდაჭერის** დონე, რომელიც გულისხმობს ენის მოხმარების უფრო დახვეწილ ფორმებს, როგორიცაა, მაგალითად, ტექსტის სორტირება სხვადასხვა პარამეტრების მიხედვით (მაგ., სიტყვათა ავტომატური დალაგება ანბანის მიხედვით), დროისა და თარიღების სწორი გენერირება

(შდრ., მაგალითად, დათარიღების ინგლისური და გერმანული სტანდარტი), ტექსტის გაწყობის ავტომატიზაცია (მაგალითად, ავტომატური დამარცვლა და ავტომატური გადატანის წესები), ტექსტების კონვერტირება და ა.შ.

ტექნოლოგიური მხარდაჭერის ქვედაზე მაღალ საფეხურს **ენის სამომხმარებლო დონეზე მხარდაჭერა** წარმოადგენს. მხარდაჭერის ეს ფორმა ქვედაზე სრულყოფილია და გულისხმობს მომსახურების სრული პაკეტის არსებობას მოცემულ ენაზე. გარდა ენის ტექნოლოგიური მხარდაჭერის ზემოთ ჩამოთვლილი კომპონენტებისა, მოცემულ შემთხვევაში უნიკოდი მომხმარებელს სთავაზობს მოცემულ ენაზე მთლიან სამომხმარებლო ბადეს: პროგრამები და მენიუები სრულად არის ადაპტირებული, ტექნიკური ხელსაწყოების ინტერაქტიულობის ხარისხი მაღალია, რაც მიიღწევა სა-განგებო ფუნქციის text-to-speech და speech-to-text მეშვეობით და აადვილებს ხელსაწყოების მოხმარებას. ასევე გათვალისწინებულია წერის მიმართულების ცვლის შესაძლებლობები: Right–to–left (RTL) languages support.

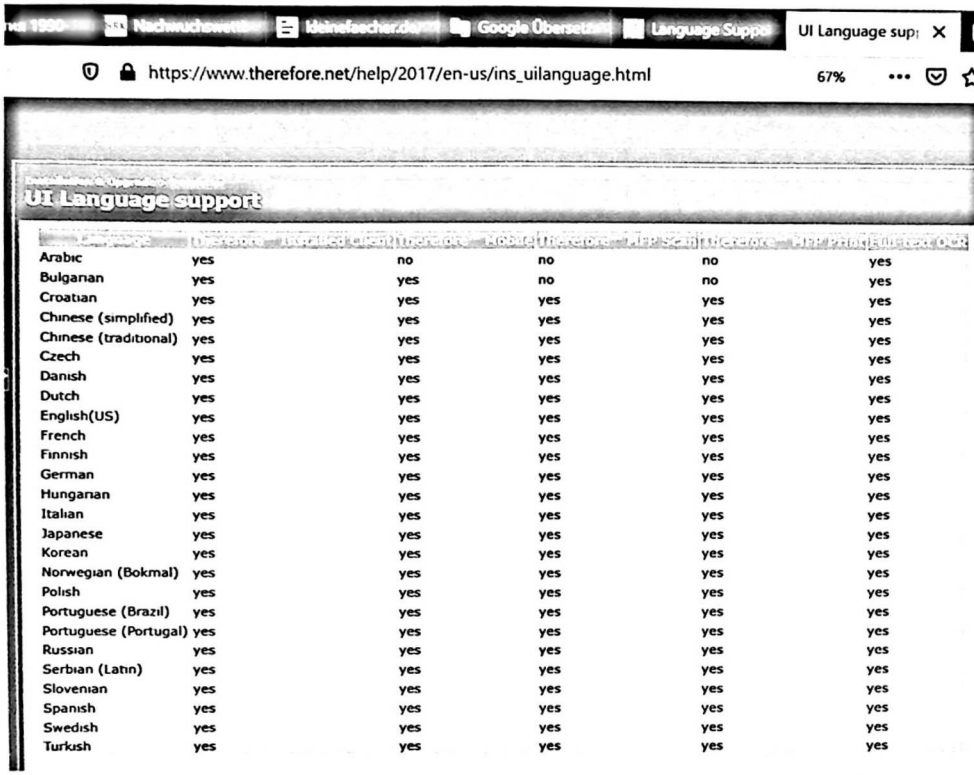
მაგალითად: მარჯვნიდან მარცხნივ წერის მხარდაჭერის მაგალითი არაბული ენისათვის:

لغة

اللغة نسق من الإشارات والرموز، يشكل أداة من أدوات المعرفة، وتعتبر اللغة أهم وسائل التفاهم والاحتكاك بين أفراد المجتمع في جميع ميادين الحياة. وبدون اللغة يتعذر نشاط الناس المعرفي. وترتبط اللغة بالتفكير ارتباطًا وثيقًا؛ فأفكار الإنسان تصاغ دومًا في قالب لغوي، حتى في حال تفكيره الباطني. ومن خلال اللغة فقط تحصل الفكرة على وجودها الواقعي. كما ترمز اللغة إلى الأشياء المنعكسة فيها.

განსაკუთრებით მნიშვნელოვანია სხვადასხვა მიმართულების წერის ერთდროული გამოყენების პრობლემის გადაჭრა ერთი დოკუმენტის (A bidirectional text, შემოკლებით BiDi) ფარგლებში. დამწერლობის ამ შერეული ტიპის გამოყენება ძირითადად სამეცნიერო ფილოლოგიური ნაშრომებისთვის არის რელევანტური, თუმცა ბიუროკრატიული ენისთვისაც (დოკუმენტაციის წარმოება ორენოვან სახელმწიფოებში) მნიშვნელოვანია.

დღეისათვის არსებობს ენის სამომხმარებლო დონეზე მხარდაჭერის სრული პაკეტი შემდეგი ენებისათვის: არაბული, ბულგარული, ხორვატული, ჩინური, ჩეხური, დანიური, უნგრული, იტალიური, იაპონური, კორეული, პორტუგალიკური, პორტუგალიური (ბრაზილიის), რუსული, სერბული, ესპანური შვედური, თურქული და ა.შ.



Arabic	yes	no	no	no	yes
Bulgarian	yes	yes	no	no	yes
Croatian	yes	yes	yes	yes	yes
Chinese (simplified)	yes	yes	yes	yes	yes
Chinese (traditional)	yes	yes	yes	yes	yes
Czech	yes	yes	yes	yes	yes
Danish	yes	yes	yes	yes	yes
Dutch	yes	yes	yes	yes	yes
English(US)	yes	yes	yes	yes	yes
French	yes	yes	yes	yes	yes
Finnish	yes	yes	yes	yes	yes
German	yes	yes	yes	yes	yes
Hungarian	yes	yes	yes	yes	yes
Italian	yes	yes	yes	yes	yes
Japanese	yes	yes	yes	yes	yes
Korean	yes	yes	yes	yes	yes
Norwegian (Bokmal)	yes	yes	yes	yes	yes
Polish	yes	yes	yes	yes	yes
Portuguese (Brazil)	yes	yes	yes	yes	yes
Portuguese (Portugal)	yes	yes	yes	yes	yes
Russian	yes	yes	yes	yes	yes
Serbian (Latn)	yes	yes	yes	yes	yes
Slovenian	yes	yes	yes	yes	yes
Spanish	yes	yes	yes	yes	yes
Swedish	yes	yes	yes	yes	yes
Turkish	yes	yes	yes	yes	yes

ამგვარი სამომხმარებლო პაკეტი საშუალებას აძლევს მომხმარებელს თავისუფლად გამოიყენოს პროგრამები მოცემულ ენაზე სხვა ენის (მაგალითად, ინგლისურის) ცოდნის გარეშე.

კონსორციუმის ერთ-ერთ მნიშვნელოვან ამოცანას წარმოადგენს კაცობრიობის ენობრივი მემკვიდრეობის შენახვა და გადარჩენა. რათა არ დაიკარგოს საუკუნეების განმავლობაში შექმნილი არამატერიალური და მატერიალური კულტურული მემკვიდრეობის ეს უმნიშვნელოვანესი ნაწილი. ამის მისაღწევად, უპირველეს ყოვლისა, უფრო და უფრო ფართოვდება იმ ენათა რიცხვი, რომლისთვისაც უნიკოდი საგანგებო ნიშნების კოდირებით მათ საერთაშორისო მიმოქცევაში მოხმარებას უწყობს ხელს. უნიკოდის მიერ სტანდარტიზებული სისტემის შექმნა სხვადასხვა ენებისათვის მუდმივად მიმდინარეობს და განაპირობებს ახალი სამომხმარებლო ვერსიების შექმნას. ბოლო ვერსიის შესახებ დაწვრილებით შეგიძლიათ იხილოთ ნაშრომი The Unicode Consortium. *The Unicode Standard, Version 12.0.0*, (Mountain View, CA: The Unicode Consortium, 2019. ISBN 978-1-936213-22-1).[142]

[142] ინტერნეტვერსია იხ. მისამართზე: http://www.unicode.org/versions/Unicode12.0.0/

კოდირების საერთაშორისო სტანდარტის შემუშავებამდეც არსებობდა კოდირების სისტემა: თითოეული სიმბოლო (მაგალითად: ასოები, პუნქტუაციის ნიშნები, ტექნიკური სიმბოლოები, მათემატიკური ოპერაციის აღმნიშვნელი ნიშნები და ა.შ.) კოდირებული იყო ციფრების მეშვეობით, რომელიც 0-ისა და 1-ის მონაცვლეობის კომბინაციის შედეგად იქმნებოდა. კოდირების ამგვარი სისტემა, როგორც წესი, ინდივიდუალურად იქმნებოდა და განსხვავდებოდა ერთმანეთისაგან არა მხოლოდ სხვადასხვა ქვეყნებისა და ენების შემთხვევაში, არამედ ერთი ენის ფარგლებშიც: ენობრივი სიმბოლოების გადმოსაცემად ერთ შრიფტში გამოყენებული ციფრების კომბინაცია მეორე შრიფტში შესაძლოა სხვა სიმბოლოს კოდირებისთვის იყო გამოყენებული, რაც შეუძლებელს ხდიდა ტექსტის ერთი კომპიუტერიდან მეორეში ტექსტის ადეკვატურად გადატანას.

თვალსაჩინოებისათვის მოვიყვანთ ქართული ენის მაგალითს. 90-იანი წლებისათვის, როდესაც პირველად შემოვიდა მასობრივ მოხმარებაში კომპიუტერი, სხვადასხვა ინსტიტუციებში სხვადასხვა პირების მიერ იქმნებოდა „ლოკალური შრიფტები“ ქართული ენისათვის, რომლებიც, მართალია, ტექსტს ერთგვაროვნად „აჩვენებდნენ“ ერთი კომპიუტერის ეკრანზე, მაგრამ ტექსტი საგრძნობლად „ირეოდა“ სხვა კომპიუტერში მისი გადატანის ან გადაგზავნის შემთხვევაში – ტექსტის სწორი წაკითხვისათვის საჭირო იყო არა მარტო ტექსტის, არამედ შესაბამისი შრიფტის გადატანაც მეორე კომპიუტერში. თვალსაჩინოებისათვის მოვიყვანთ გიორგი გოგოლაშვილის სტატიას „ქართული, როგორც სახელმწიფო ენა“, რომელიც გამოქვეყნებულია „იბერიულ-კავკასიური ენათმეცნიერების“ XXXVII ტომში.

XIX საუკუნის 60-იან წლებში დაიწ ო ფართომასშტაბიანი ეროვნულ-განმათავისუფლებელი მოძრაობა. ამ მოძრაობის ერთ-ერთი მთავარი მიზანი ი ო საქართველოში ქართული ენისათვის თავისი კუთვნილი ადგილის დაბრუნება; რუსული ენის მიერ წართმეული ფუნქციონირების სფეროს ა დაგენა. ეროვნულ-განმათავისუფლებელი ბრძოლის ამ ეტაპის დასაწ ისად ითვლება ილია ჭავჭავაძის წერილის – „ორიოდე სიტ ვა თავად რევაზ შალვას ძის ერისთავის მიერ კოზლოვიდგან „შეშლილის“ თარგმანზედ“ – გამოქვე ნება 1861 წელს. რა ი ო ილია ჭავჭავაძის მთავარი თეზისი: „სამი ვთაებრივი საუნჯე დაგვრჩა ჩვენ მამა-პაპათაგან: მამული, ენა და სარწმუნოება. თუ ამათაც არ უპატრონებთ, რა კაცები ვიქნებით, რა პასუხს გაეცემთ შთამომავლობას?!“ (ილია 1958, გვ. 26).

როგორც ვხედავთ, რამდენიმე გრაფემა – ყ და ღ არ იკითხება და შესაბამისი ადგილი ცარიელია. ასო ყ რეალიზდება ასო ჭ-ს მაგიერად (ყაენყავაძე < ჭაენჭავაძე),

ხოლო ასო ჭ რეალიზდება ჯ-ს მაგიერად და პირიქით, ასო ჯ-ს მაგიერად ჭ რეალიზდება:

> რამდენადაც უცნაურად უნდა ჯ ერდეს, საბჭოთა ხელისუფლება ქართულზე, როგორც სახელმწიფო ენაზე, აქტიურ ზრუნვას განაგრძობს:

გაცილებით დრამატულია სურათი მომდევნო ფრაგმენტის შემთხვევაში – ტექსტში გამოყენებულია ინდივიდუალურად შექმნილი შრიფტი Kartveluri_Norm, რომელიც არათავსებადია საერთაშორისო სტანდარტთან. მაგალითად, ლელა გიგლემიანის სტატია „რედუპლიკაციით მიღებული ზმნისართები სვანურში“, რომელიც „იბერიულ-კავკასიური ენათმეცნიერების“ XXXVIII ტომში არის გამოქვეყნებული, ჩემს პირად კომპიუტერში არ იკითხება, ვინაიდან ტექსტი აკრეფილია შრიფტით Kartveluri_Norm, რომელიც სრულიად არათავსებადია ჩემს კომპიუტერში ჩაწერილ შრიფტებთან:

> ÊÏÌÐÏÆÉÝÉÀ(ÀÍÖ ÈáÆÅÀ)ÀáÀË ËÄØÓÉÊÖÒ ÄÒÈÄÖËÈÀ ßÀÒÌÏÓÀØÌÍÄËÀÃ ÂÀÌÏÚÄÍÄÁÖËÉÀ ÌÒÀÅÀË ÓáÅÀÃÀÒÓáÅÀ ÓÉÓÔÄÌÉÓ ÄÍÀÙÉ, ÌÀÈ ÛÏÒÉÓ _ ØÀÒÈÅÄËÖÒ ÄÍÄÁÛÉÝ. ÄÓ ÄáÄÁÀ ÊÏÌÐÏÆÉÝÉÉÓ ÄÒÈ-ÄÒÈ ÓÀáÄÏÁÀÓÀÝ _ ÒÄÃÖÐÐËÉÊÀÝÉÀÓ, ÒÏÏÄËÉÝ ÀáÀËÉ ÓÉÔÚÅÄÁÉÓ ÌÉÙÄÁÀÓ ÂÖËÉÓáÌÏÁÓ ÄÒÈÉ ÃÀ ÉÂÉÅÄ ×ÖÞÄÈÀ ÂÀÍÌÄÏÒÄÁÉÈ, ÂÀÏÒÊÄÝÄÁÉÈ.

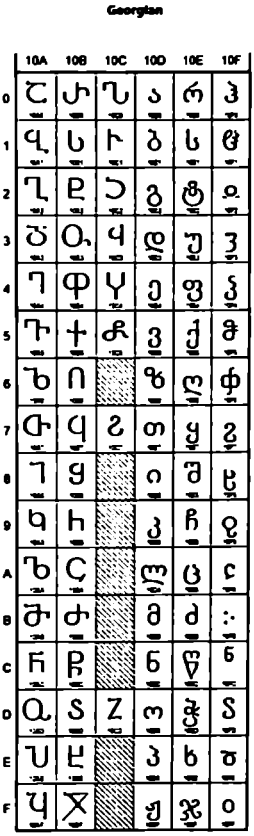
ტექსტის წაკითხვის მცდელობა შრიფტით AcadNusx Unicode სრულიად უშედეგოა, ვინაიდან ამ შემთხვევაში ტექსტი ასე გამოიყურება:

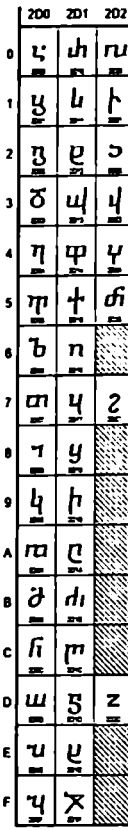
> Θ(Θ Θ)Θ Θ ϑ ϑ ϑ Θ Θ Θ ϑ ΘϚ ϰΘ ϑ ϑα -
> Θ Θ Θ ΘϚΘ Θ ϑ ϑ Θ , Θ – Θ ϑ ϑ ϑα . ϑ ϑ ϑαΘ
> ϑ -ϑ Θ ϑ αΘ Θ – ϑϚ Θ Θ , ϑ Θ Θ ϑα ϑ-
> αΘ ϰ α ϑ ϚΘ ϰ ϑ × ϑ Θ ϰΘ ϑ ϑα , ϰΘ ϑ ϑα .

შესაბამისად, იმისათვის, რომ ტექსტი წავიკითხოთ, საჭიროა მისი კონვერტირება, მაგალითად Sylfaen-ში, რომლის შედეგადაც ტექსტი ასე გამოიყურება:

> კომპოზიცია (ანუ თხზვა) ახალ ლექსიკურ ერთეულთა წარმოსაქმნელად გამოყენებულია მრავალ სხვადასხვა სისტემის ენაში, მათ შორის – ქართველურ ენებშიც. ეს ეხება კომპოზიციის ერთ-ერთ სახეობასაც – რედუპლიკაციას, რომელიც ახალი სიტყვების მიღებას გულისხმობს ერთი და იგივე ფუძეთა განმეორებით, გაორკეცებით.

კოდირების ახალმა სტანდარტმა შესაძლებელი გახადა ეროვნული ანბანების საყოველთაო სტანდარტიზაცია, მათ შორის, ჩინური იეროგლიფებისაც. დღეისათვის ქართული ანბანის სამივე სახე არის უნიკოდში ასახული: ასომთაერული (U+10A0 – U+10C5)[143], ნუსხა-ხუცური (U+2D00 – U+2D25)[144] და მხედრული (U+10D0 – U+10F5).[145]



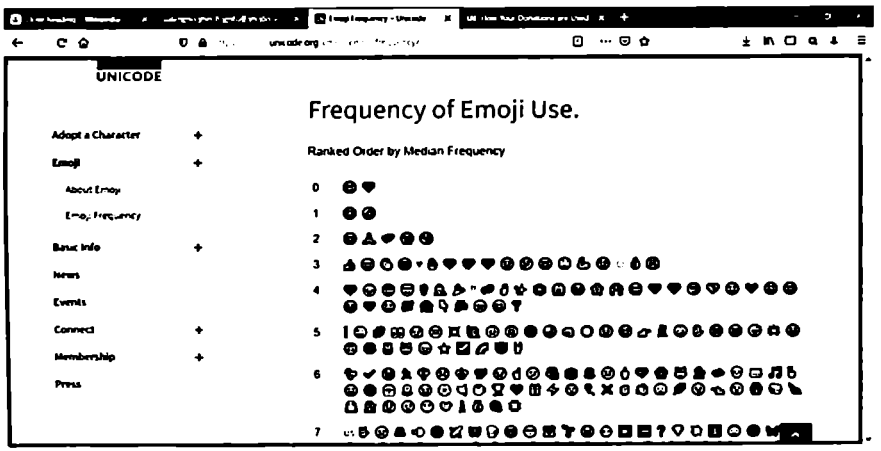


საგულისხმოა ის გარემოება, რომ უნიკოდის კონსორციუმი არა მარტო შრიფტების საერთაშორისო სტანდარტიზაციით არის დაინტერესებული, არამედ დიგიტალური კომუნიკაციის სხვა ნიშნების სტანდარტიზაციითაც. გლობალური საინტერნე-

[143] http://www.unicode.org/charts/PDF/U10A0.pdf
[144] http://www.unicode.org/charts/PDF/U2D00.pdf
[145] ზოგადი ინფორმაცია UNICODE-ის შესახებ შეგიძლიათ იხილოთ მისამართზე: http://www.unicode.org, ხოლო UNICODE 7-ის ფარგლებში შემუშავებული კოდები http://www.unicode.org/charts/

ტო სივრცის შექმნამ და სწრაფმა განვითარებამ ინფორმაციის გადაცემის ეკონომიურობის აუცილებლობა გამოიწვია, რამაც, თავის მხრივ, ახალი სემიოტიკური ნიშნების – **ხატულების** (icon) შექმნა განაპირობა. დღეისათვის საერთაშორისო მიმოქცევაშია ხატულების დიდი რაოდენობა, რომელთა გამოყენება და გაშიფვრა იმავე პრობლემებთან არის დაკავშირებული, როგორიც შრიფტების შემთხვევაში იყო: სხვადასხვა ქვეყანაში სხვადასხვა დროს შექმნილი ხატულების „წაკითხვა“ ყოველთვის არ არის შესაძლებელი, ვინაიდან მას საერთაშორისო პროგრამული უზრუნველყოფა არ გააჩნია. სწორედ ამიტომ დაიწყო უნიკოდმა ხატულების სტანდარტიზაცია და შექმნა ხატულების ბაზა, რომელიც მოიცავს როგორც ხატულების ჩამონათვალს, ისე მის ვერბალურ შინაარსსაც. თითოეულ ხატულას თავისი უნიკალური კოდი აქვს მინიჭებული, რაც უზრუნველყოფს საერთაშორისო მიმოქცევაში მყოფი ხატულების უპრობლემო „წაკითხვადობას“.

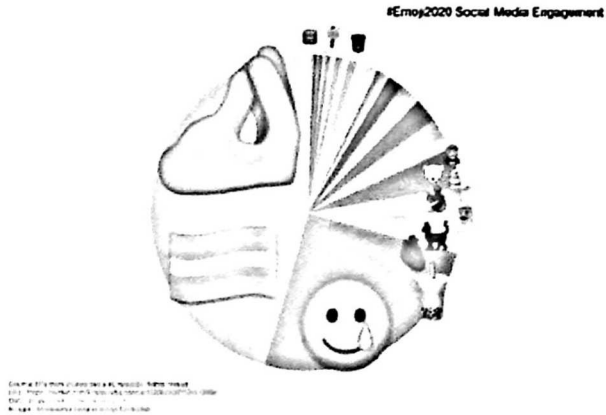
ქვემოთ მოცემულია 2019 წელს კოდირებული ახალი ხატულები და მათი მოხმარების სიხშირე სოციალურ ქსელებში. უნიკოდის გვერდზე შეგიძლიათ იხილოთ ყველაზე ხშირად გამოყენებული ხატულების რეგისტრი.



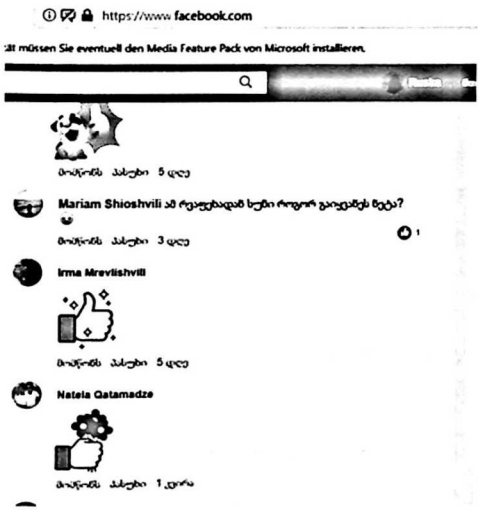
სიმბოლოების სია ყოველწლიურად ახლდება უნიკოდის კონსორციუმის მიერ. 2020 წლის 10 თებერვლის მონაცემებით ყველაზე ხშირად გამოყენებული ხატულებია: შეერთებული თითები (შეესაბამება იტალიურ ჟესტიკულაციას მნიშვნელობით *mia che vuoi „რა გინდა?“*), ტრანსგენდერთა დროშა და მოცინარი სახე ცრემლებით (Pinched Fingers, Transgender Flag, Smiling Face With Tear):

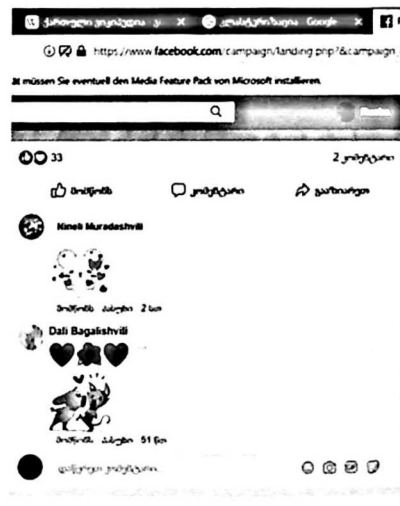


2020 წელს ყველაზე ხშირად ხმარებული ხატულების სრული სურათი მოცემულია ქვემოთ:



ქართულ მედიასივრცეში სულ უფრო მატულობს ხატულების გამოყენების პრაქტიკა. ხშირია ხატულების კომბინირებული გამოყენების შემთხვევები:





საკითხი იმის თაობაზე, რა ვერბალური შინაარსის მატარებლები არიან მოყვანილ მაგალითებში მოცემული ხატულები ცალ-ცალკე, ან როგორ უნდა „წაიკითხოთ“ ისინი ხატულების კომბინაციის შემთხვევაში, ცალკე კვლევის საკითხია და ქართული ენის არავერბალური ფორმებით გადმოცემის თვალსაზრისით უაღრესად მნიშვნელოვან კვლევით ასპექტს წარმოადგენს.

1.1. კოდირების სპეციალური პრობლემები ქართული ენის ეროვნულ კორპუსში: ინიციალების, ქარაგმების, ფერების, ზომის კოდირება

ვერიფიკაციის მნიშვნელობიდან გამომდინარე ქართული ენის ეროვნულ კორპუსში ერთ-ერთ მნიშვნელოვან ასპექტს წარმოადგენს ინიციალების, ქარაგმების, ფერებისა და ზომების კოდირება. კოდირებაში ამ შემთხვევაში იგულისხმება „ნორმალიზაციის“ საკითხი. ტექსტების კოდირების დროს, როგორც წესი, დაცულია დედანში გამოყენებული ნიშნები, თუმცა ასომთაერულით შესრულებული ტექსტები კორპუსში გადმოცემულია როგორც ასომთავრულით, ისე მხედრულით. ამასთან დიგრაფემები, რომელიც ასომთავრულისთვის არის დამახასიათებელი (მაგ., უ-ს გადმოცემა Ⴓ-ს და Ⴕ-ს კომბინაციით), მხედრულით გადმოცემის დროსაც ინარჩუნებს ფონემის დიგრაფემული გადმოცემის პრინციპს. მაგ.:

- ასომთაერული → მხედრული
  - ⴑⴓⴕⴆⴎⴆⴏⴕⴇ, ⴆⴑ ⴑⴏⴓⴕⴆⴑⴎⴇ
  - სოულნელები, და სხოუანი

გრაფემების ნორმალიზაციის პრინციპი ასევე დაცულია ნუსხურით შესრულებული ტექსტების მხედრულით გადმოცემის შემთხვევაშიც:

- ნუსხური → მხედრული
  - *ⴆⴑ ⵗⴅⴎⵗⴅⴍⵗⴆⴅⴍⴆⴎⴆⵗ ⴝⴏⴅⴎⴍⴎⵗⴅ ⴆⴅⴍⴅⴅⴅ:*
  - და გურგუნ-შემოსილი იხარებს ზეცას.

ასევე დაცულია საზედაო ასოების გადმოცემის სიზუსტე როგორც ასომთავრულ, ისე შერეულ (ასომთავრულით და ნუსხურით შესრულებულ) ტექსტებში.

- ⴆⴑ ⴚⴓⴏⴑⴕⴈⴓⴎⴆⴑⴆⴑ → ⴆა მოხაქოუნდა
- *ⴍⴅⴎⴏⴑⴚⴆⴅⴅ ⴓⴅⴎⴆⵗⴅⴎⴎⴆⴅⴅ* → ⴍესხმითა სულიერითა

განსაკუთრებულ პრობლემას ქართული ენის ტექსტების კოდირების პროცესში ქმნის ქარაგმების საკითხი.[146] ტერმინი ქარაგმა ბერძნულიდან შემოსული სიტყვაა და ნიშნავდა შუბლის ან ხელის ნაოჭს. სულხან-საბა ორბელიანმა მის აღსანიშნად ქართული ტერმინი შექმნა და მას „ასონაკლების ნიშანი“ უწოდა, თუმცა ქართული ტერმინი ვერ დამკვიდრდა და სამეცნიერო მიმოქცევაში დღესაც ტერმინი ქარაგმა იხმარება. ტერმინი პირველად მე-10 საუკუნეში, ცნობილი ჰიმნოგრაფის, მიქაელ მოდრეკილის მიერ შედგენილ საგალობელთა კრებულში „იადგარშია“ (ხელნაწერი S 425, 978-988 წწ.) ნახმარი, ძლისპირების ბოლოს დართულ მიქეილ მოდრეკილის ანდერძში. თუმცა მისი გამოყენება – ასოთა გამოკლების წესი, უძველეს ხელნაწერებში და წარწერებში დასტურდება.

დაქარაგმების წესებს საგანგებოდ შეეხო ანტონ I კათალიკოსი თაეის „ქართულ ღრამმატიკაში“. მან ამ ნაშრომის პირველ რედაქციაში (1753) გამოყო ოთხგვარი ქარაგმა:

1. საუფლო (ღ̃ი – ღმერთი, ო̃ი – უფალი, ი̃ჯ – იესუ, ქ̃ე – ქრისტე, წ̃ჲ – წმიდაჲ, ან̃გს – ანგელოს, იი̃ლ – ისრაელ, იი̃ლიმ – იერუსალიმ, კ̃სი – კათალიკოსი); (nomina sacra)

2. სამთავრულო (ხ̃ – ხოლო, რ̃ – რამეთუ); (suspension)

3. საყოველადო (ა̃დ – არამედ, ვ̃დ – ვითარმედ, ე̃ე – ეიდრე, ყ̃ი – ყოველი, რ̃ლი – რომელი, კ̃დ – კუალად, შ̃ს – შორის...);(contraction)

4. საჭმოუანო (ჩ̃მი – ჩემი, შ̃ნი – შენი...).

თუმცა ნაშრომის მეორე რედაქციაში იგი ამარტივებს ამ კლასიფიკაციას და ქარაგმის ორგვარ სახეს გამოყოფს: უაღრესობითი (საყოველადო+სამთავრულო) და საყოველადოო (საყოველადო + სამთავრულო + საჭმოუანო)

ივ. ჯავახიშვილმა, რომელმაც საფუძვლიანად შეისწავლა ძველ ქართულში სიტყვათა შემოკლების წესები, მოახდინა მათი სხვაგვარი სისტემატიზაცია. ასონაკლების პოზიციისა და რაოდენობის მიხედვით მან ქარაგმა შემდეგ ქვეტიპებად დაყო: დაქარაგმება (ზოგადი ტერმინი), ბოლოშეკეცილობითი დაქარაგმება (suspension), შუ-

[146] ქართულ ძეგლებში დაქარაგმების საკითხი საგანგებოდ იქნა შესწავლილი ფრანკფურტის უნივერსიტეტის პროექტში LOEWE. ამ საკითხის განხილვისას ჩვენ მოღიანად ეყრდნობით პროექტში მონაწილე ჩვენი კოლეგების ლელა სამუშიას, ნათია დუნდუასა და არმინს პოენინის მიერ შეკრებილ მასალას და დასკვნებს. იხ.: დუნდუა ნ., პოენინი ა., სამუშია ლ., ძველი ქართული სახარების ხელნაწერების პარალელური კორპუსი და სტემატოლოგია, "ენა, ლოგიკა, კომპიუტერი", 2015.

აშკარაცილობითი (contraction), მხოლოასოიანი (მხოლოდ დასაწყისი ასოთი წარმოდგენილი სიტყვისათვის), უხმოენო (უხმოვნოდ დაწერილი სიტყვებისათვის).

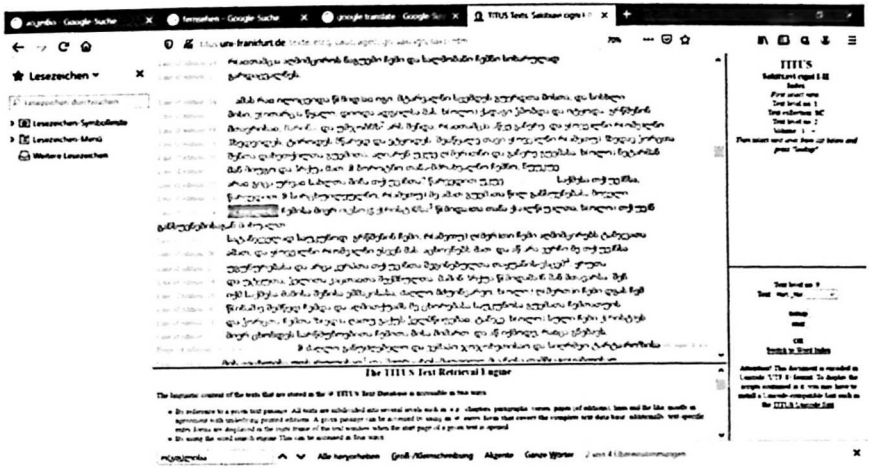
ძველ ქართულ ხელნაწერებში დაქარაგმების რაიმე განსაზღვრული მეთოდი არ არსებობდა, თუმცა მაინც შეიძლება გარკვეული სისტემის დანახვა. როგორც წესი, ქარაგმდებოდა "წმინდა სახელები" (Nomina sacra), ვინაიდან ბერძნულში, საიდანაც ითარგმნებოდა უამრავი ქართული ტექსტი, მხოლოდ "წმინდა სახელები" იყო დაქარაგმებული და ასეთი სახელები ქართულ თარგმანშიც ისევე იყო გადმოღებული, როგორც ბერძნულ წყაროში. თუმცა, მათ გვერდით ქართულში სხვა სახელებიც მოკლდებოდა, კერძოდ კავშირები, ზმნისწედები, თანდებულები, ზმნისწინებიც კი.

აღსანიშნავია, რომ დასაქარაგმებელი სიტყვების რიცხვი თანდათანობით იზრდება მომდევნო ხანის ხელნაწერებში. მაგ., იერუსალიმის საბაწმიდის ლავრაში 864 წელს გადაწერილ ე.წ. სინ'ურ მრავალთავში საგრძნობლად გაზრდილია დაქარაგმებული სიტყვების რიცხვი. ჩანს, უხმოენო დაქარაგმება მე-9 საუკუნიდან ფართოდ იკაფავს გზას, რადგან საყრდენი თანხმოვნებითა და გრამატიკული ღირებულების მქონე ხმოვნიანი დაბოლოებით დაწერილი სიტყვები წაკითხვის თვალსაზრისით დიდ სირთულეს არ ქმნიდა.

დაქარაგმებას ქართულში წმინდა პრაქტიკული მნიშვნელობა ჰქონდა: საწერი მასალის ეკონომია. ამის საუკეთესო დამამტკიცებელი ფაქტი ის არის, რომ ძველ ქართულ ხელნაწერებში შემოკლებული სიტყვების ყველაზე დიდი რაოდენობა სტრიქონის ბოლოს მოდიოდა. ხელნაწერებიდან გამოიყოფა რიგი სიტყვებისა, რომლებსაც გადამწერები ჩვეულებრივ არ აქარაგმებენ, მაგრამ სტრიქონის ბოლოს უადგილობის გამო შემოკლებულად წერენ.

უძველეს ხელნაწერ ტექსტებში, V-VIII საუკუნეთა ხანმეტ და ჰაემეტ ფრაგმენტებში, თითქმის ერთგვარად არის დაქარაგმებული ისეთი "წმინდა სახელები", როგორიცაა ი҃ჯ (იესუ), ქ҃ე (ქრისტე), ო҃ი (უფალი), ღ҃ი (ღმერთი), ჯ҃რი (ჯუარი), დ҃თ (დავით), ფს҃ლნი (ფსალმუნი), იშ҃ლი (ისრაჱლი), იშ҃მი (იერუსალჱმი). გარდა ამისა, დაქარაგმებულია: 1. კავშირები: რ҃ (რამეთუ), ხ҃ (ხოლო), ა҃დ (არამედ), ვ҃ე (ვიდრე), ვ҃რ (ვითარ), ვ҃ა (ვითარცა), ე҃დ (ვითარმედ), რ҃ა (რაჲთა); 2. ზმნისწედები: ფ҃დ (ფრიად), ა҃ნ (ამენ), რუ҃(მ)ს (რაჟამს), მ҃ს (მარადის), ო҃კე (უკუნისამდე); 3. თანდებულები: ზ҃ა (ზედა), წ҃ე (წინაშე), გ҃ნ (გან); 4. ნაცვალსახელები: ყ҃ლი (ყოველი), რ҃ლი (რომელი), ჩ҃ნ (ჩუენ), თქ҃ნ (თქუენ); 5. ზედსართავი სახელები: ეგეე҃რი (ეგეეითარი), ესეე҃რი (ესეეითარი), არაწ҃ჲ (არაწმიდაჲ), წ҃ჲ (წმიდაჲ); 6. ზმნები: თქ҃ა (თქუა), ხრ҃ქ҃ა/ხრ҃ქ҃ა (ხრქუა/ხრქუა)...

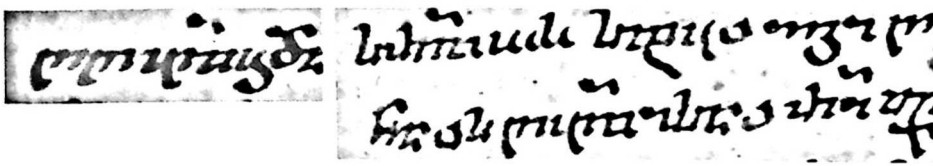
TITUS-ის პლატფორმაზე განთავსებულ ხელნაწერთა ტექსტების ციფრულ ვერსიებში ქარაგმები გახსნილია და საგანგებო ნიშნებითაა მარკირებული:



მაგ.: ⵁⵛⵒⵏⵙⵛ > ო(ჳფ)ლისა, ⵜⵗ > ი(ესო)ჳ, ⵋⵒⵏⵙ ქ(რისტ)ჱსა

ⵒⵒⵒⵒⵒⵒ ⵏⵒⵒⵒⵒ ⁖ → ცხ(ოვ)რ(ე)ბ(ა)სა ს(აჳ)კ(უ)ნესა

დაქარაგმების პრინციპები ქართული ენის უძველეს ძეგლებში არც ცალსახაა და არც ერთგვაროვანი. ქართულ ხელნაწერებში დაქარაგმების პრინციპების შესწავლამ გვიჩვენა, რომ ხელნაწერები დაქარაგმების თვალსაზრისით არსებით განსხვავებას გვიჩვენებენ. მაგალითად, სიტყვა წინასწარმეტყველება სხვადასხვა ფორმით არის დაქარაგმებული ხელნაწერებში. შესაბამისად, პრობლემურია ამგვარი დაქარაგმების შემთხვევაში ქარაგმების გახსნა.



განსაკუთრებული ვითარებაა ე.წ. "ჭილ-ეტრატის იადგარში" (H-2123), რომელიც პალესტინის საბაწმიდის ლავრაშია გადაწერილი X საუკუნეში. ამ ხელნაწერში თითქმის ყველა სიტყვა ქარაგმდება. როგორც არ. მარტიროსოვი მიუთითებს, მაგალითად, ერთ-ერთ გვერდზე სულ 73 სიტყვაა ნახმარი, აქედან 67 ქარაგმიანია, 6

– უქარაგმო. ეს თავისებურება გამოწვეული უნდა ყოფილიყო როგორც ძნელად საშოვნელი საწერი მასალის – ჭილისა და ეტრატის სიმცირით, ისე თვით საგალობელთა შინაარსით, რომელიც, სავარაუდოდ, სათანადო პირებმა მეტწილად ზეპირად იცოდნენ. ამასთან, გასათვალისწინებელია ისიც, რომ ერთი და იგივე სიტყვა ყოველთვის ერთნაირად არ არის შემოკლებული.

აღსანიშნავია, რომ ხელნაწერებში დაქარაგმების სხვადასხვა წესი დასტურდება:

1. ერთი ქარაგმის ქვეშ ხშირად შემოკლებულია ორი დამოუკიდებელი ან ერთი რთული სიტყვა: ქს̃მრთ (ქრისტჱს მიმართ), ქლ̃მდი (ქველისმოქმედი), მყლ̃რს (მომაყალ არს), დბლ̃ხრ (დიდებულ ხარ)...

2. ერთი ქარაგმის ქვეშ შემოკლებით წარმოდგენილია მთელი წინადადება და მისი წაკითხვა მეტად გაძნელებულია: მხგყ̃ნნ (მეოხ გუეყავ ჩუენ), მეოხებითა შთა̃აცხსი (მეოხებითა შენითა აცხოვნენ სულნი ჩუენნი), შჳყნენ ს̃ლნი (შეიწყალნენ სულნი ჩუენნი).

3. შემოკლებისას ზოგჯერ არაა წარმოდგენილი სიტყვის ბოლო ნაწილი, ბრუნვის და რიცხვის ნიშნები, ზმნური სუფიქსები (ან მათი ნაწილი): აღვიესენით სხ̃ლ (აღვიესენით სიხარულითა), ცხრ̃ბდ სლ̃ ჩ̃ (ცხორებად სულთა ჩუენთათჳს), აკრ̃თ (აკურთხეუდით), აც̃ ნი̃ (აცხოვნენ სულნი ჩუენნი), მწ̃თა სხ̃რლ (მწეთა სიხარულო) და სხვ.

4. თანდებული გან და ზმნისწინი გან, მოთხრობითი ბრუნვის მან ნიშანი ხშირად მხოლოდ დასაწყისი თანხმოვნით არიან წარმოდგენილი: შ̃ნგ (შენგან), რი̃საგ (რომლისგან), საცთურისაგ̃ (საცთურისაგან), სიმჴნემ̃ შნ̃მ (სიმჴნემან შენმან) და სხვ.



პრობლემურია აგრეთვე სასვენი ნიშნებისა და ასოების რიცხვითი მნიშვნელობებით გამოყენების შემთხვევების კოდირება კორპუსში, ვინაიდან უნიკოდში დაქარაგმების აღნიშნული ნიშნები ჯერჯერობით არ არის ასახული.

მაგ.: ᲝᲢ : ᲨᲔᲛᲓᲒᲝᲛᲘᲡ : Ბ : ᲥᲘᲡᲔ : ᲪᲮᲪᲮᲘᲔᲡᲘ → თ(თუეს)ა : დეკემბერსა : ა : წ(მიდ)ისა : ანანიასი → თთუესა დეკემბერსა 1 წმიდისა ანანიასი

როგორც ზემოთ მოყვანილი პრობლემის დახასიათებიდან ჩანს, ტექსტების კორპუსში ადეკვატური წარმოდგენა უკავშირდება არა მარტო ასოების კოდირების საკითხს ("character encoding"), არამედ ტექსტის შიდა მარკირების საგანგებო აღნიშვნის მანქანურ შესაძლებლობებსაც ("text markup").

1.2. ნორმალიზაციის თანამედროვე მიდგომები

ტრადიციული ლინგვისტიკა კარგად იცნობს ტექსტის კრიტიკული რედაქციის ფენომენს. რომელიმე ძეგლის გამოცემის შემთხვევაში ერთმანეთს უდარდება სხვადასხვა რედაქციები და იქმნება ერთი, კრიტიკული რედაქცია.

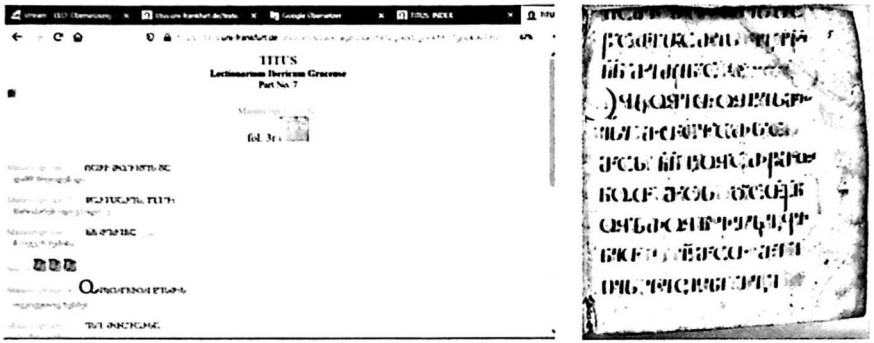
კორპუსში ტექსტების კოდირების სიზუსტის აღსაწერად ანოტაციის ორი დონე გამოიყენება:

- დიპლომატიური დონე (ნაბეჭდის თუ ხელნაწერი ტექსტის ზუსტი ეკვივალენტი, incl. ორთოგრაფიული და გრამატიკული შეცდომები და ა.შ.)
- ნორმალიზებული (ორთოგრაფიისა და გრამატიკის მიხედვით გადამუშავებული – "კრიტიკული გამოცემები", რომელიც ისტორიული ეტაპებისდა შესაბამისად არის გამართული)

ანუ, ტექსტების ელექტრონული ვერსიების შემთხვევაში ერთმანეთისაგან განასხვავებენ დედნის ზუსტ, ე.წ. **დიპლომატიურ ვერსიას** და ნორმალიზებულ, ანუ **კრიტიკულ ვერსიას**. დიპლომატიური ვერსია წარმოადგენს ხელნაწერის ზუსტ ასლს, რომელშიც უპირობოდაა შესული ყველა გრაფემა, ყველა დამატება და შესწორება (ცდომილების ჩათვლით). კრიტიკულ ვერსიაში კი მეცნიერის (ან მეცნიერთა ჯგუფის) მიერ შერჩეული თეორიული ჩარჩოს შესაბამისად ხორციელდება ტექსტის ნორმალიზაცია. შესაბამისად, ანოტირებულ კორპუსებში ინტეგრირებული ელექტრონული ტექსტები მოიცავენ მრავალდონიანი ანოტაციის სისტემას, რომელშიც სხვა მახასიათებელთა გვერდით ტექსტის ნორმალიზებული ვერსიაც არის წარმოდგენილი (ორთოგრაფიული ნორმალიზება).

1.3. ვერიფიკაციის საკითხი

ელექტრონული ტექსტის ნორმალიზაციისათვის არსებითი მნიშვნელობა აქვს ვერიფიკაციის საკითხს. იმ კორპუსებისათვის, რომლებიც ხელნაწერი ძეგლების ელექტრონულ ვერსიებს მოიცავენ, ვერიფიკაციის ყველაზე დახვეწილი ფორმა არის ფოტოპირების დართვა.



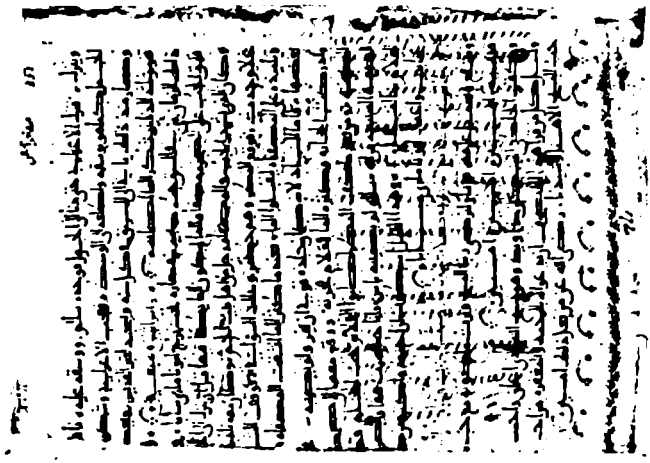
გრაცის ლექციონარის დიგიტალური ვერსია (3r) ფოტოპირითურთ

როგორც წარმოდგენილი ფოტომასალიდან ჩანს, TITUS-ში განთავსებული ქართული ხელნაწერების უმრავლესობა ვერიფიცირებადია – კოდირების საერთაშორისო სტანდარტის მიხედვით კონვერტირებულ ელექტრონულ ტექსტს თან ახლავს ხელნაწერის შესაბამისი გვერდის ფოტოასლი, რომელიც საშუალებას გვაძლევს კორპუსში ჩართული ლინგვისტური ტექსტი შევუდაროთ ორიგინალს და დავრწმუნდეთ მის ადეკვატურობაში, ან თავად მოვახდინოთ ტექსტში გაუშიფრავი ადგილების გაშიფვრა. ვერიფიკაციის ამ ფორმას განსაკუთრებული მნიშვნელობაც აქვს ე.წ. „ვირტუალური ბიბლიოთეკის/საცავის“ კონტექსტში; იგი საშუალებას აძლევს მეცნიერს სახლიდან გაუსვლელად იმუშაოს სხვადასხვა ქვეყნებში გაბნეულ ხელნაწერებზე.

2. ხელნაწერების დიგიტალიზაცია

წერილობითი ძეგლების (წარწერებისა და ხელნაწერების) დიგიტალიზაცია ხშირად დიდ სირთულეებთან არის დაკავშირებული, ვინაიდან საუკუნეებგამოვლილი ძეგლები ხშირად დაზიანებული სახით არის ჩვენამდე მოღწეული. ასეთ შემთხვევაში განსაკუთრებით ძნელია პალიმფსესტების გაშიფვრა. პალიმფსესტების ტრადიციული მეთოდებით დამუშავება გაშიფვრის მიზნით თავდაპირველად გულისხმობდა ხელნაწერის ქიმიურ დამუშავებას. თუ რამდენად საზიანოა ხელნაწერისათვის აღნიშნული

მეთოდი, კარგად ჩანს სინური კოდექსის[147] ერთ-ერთი ფრაგმენტის მაგალითზე, რომელიც ლაიფციგის უნივერსიტეტის ბიბლიოთეკაში არის დაცული. როგორც ქვემოთ მოყვანილი ქიმიურად დამუშავებული ფრაგმენტიდან ჩანს, პალიმფსესტი ორენოვანია: თავდაპირველი ბერძნული ტექსტი მოგვიანებით არაბული ტექსტით იქნა „გადაფარული“:



Cod.gr.2, fol. 22b

მე-19 საუკუნეში გამოყენებული ეს მეთოდი მე-20 საუკუნეში შეცვალა ფოტოგრაფირების მეთოდმა, კერძოდ, ულტრაიისფერი და ინფრაწითელი ფოტოების გამოყენებამ შესაძლებელი გახადა ქვედა ფენის კონტრასტირება ზედა ფენასთან, რამაც არსებითად გადაარჩინა პალიმფსესტური ხელნაწერები ფიზიკურ განადგურებას.

პალიმფსესტების დამუშავების აღნიშნული მეთოდი დღეისათვის მეცნიერების განელილი ეტაპია. თანამედროვე მეცნიერება დღეს გაცილებით უფრო კომპლექსურ ტექნოლოგიებსა და კვლევის მეთოდებს იყენებს, ვიდრე ტრადიციული მეთოდებია. ქართული (და არამარტო ქართული) ხელნაწერების შემთხვევაში განსაკუთრებით ეფექტური აღმოჩნდა პალიმფსესტების კვლევა მულტისპექტრული აპარატის MUSIS-ის გამოყენებით. 21-ე საუკუნეში შექმნილი ხელნაწერთა დამუშავების ეს ტექნოლოგიური სიახლე საშუალებას იძლევა ხელნაწერი, მასზე შესრულებული საწერი საღებავების სხვადასხვა დონეების გათვალისწინებით, განსხვავებულ დონეებად დაშა-

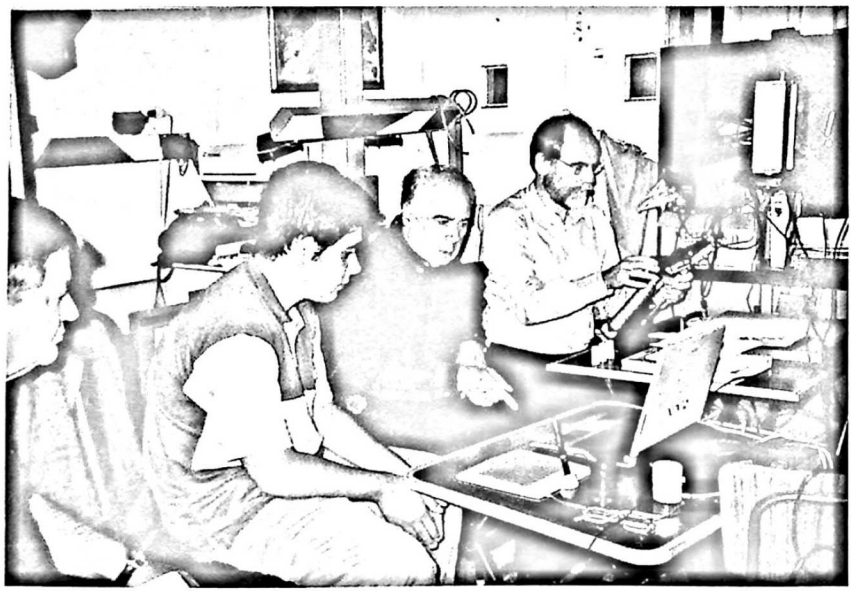
[147] იხ. Codex Sinaiticus, http://www.codex-sinaiticus.net/

ლოს და დააცალცალკევოს. ზედა ფენის ნაწერის "გაფერმკრთალება" გაცილებით "ხილულს" ხდის ძველ, ქვედა ფენას და აადვილებს ხელნაწერის ტექსტის გაშიფვრას. MUSIS-ის აპარატის გამოყენებით ხორციელდება ფერების გამის სრული სპექტრის გამოყენება ულტრაიისფერიდან ინფრაწითელამდე:

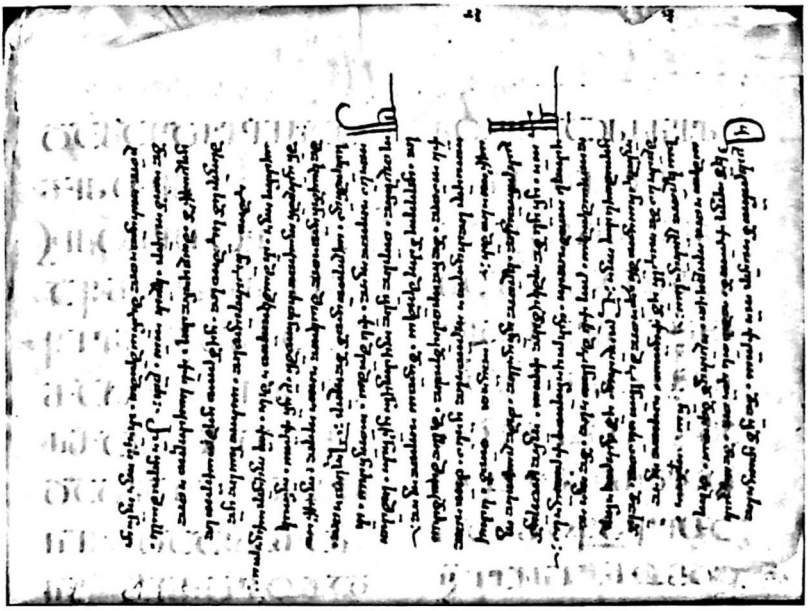
UV – violet – blue – green – yellow – orange – red – IR

…370 – 440 – 480 – 540 – 580 – 620 – 680 – 760

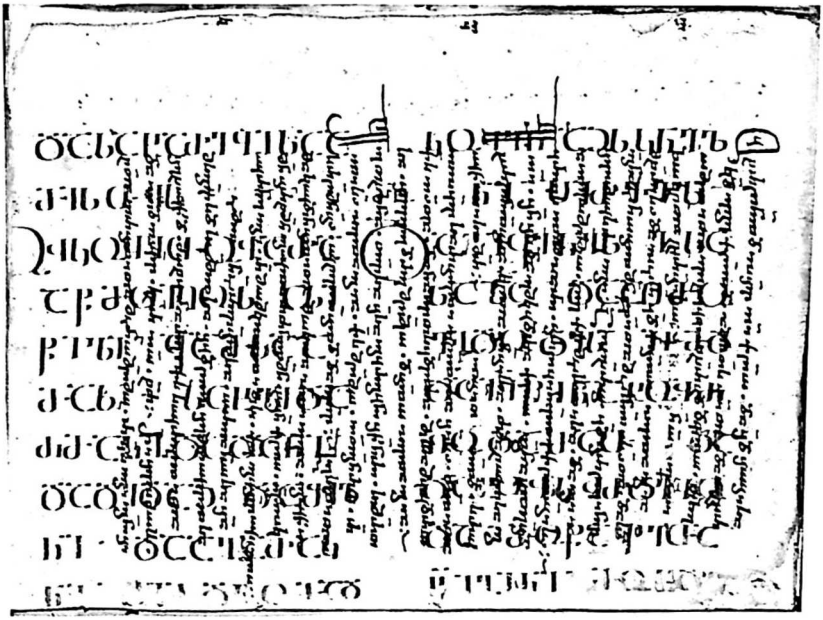
ფერების გამის სრული სპექტრის ამგვარი გამოყენება და გადაღებული ფოტოების „ზედდება" მაქსიმალურად ზრდის ქვედა ფენის ხილვადობას. უახლესი დიგიტალური მეთოდებით გაიშიფრა სინას მთის ქართული და ალბანური ხელნაწერები და ვენის პალიმფსესტები (**Cod.Vind.georg. 2**).



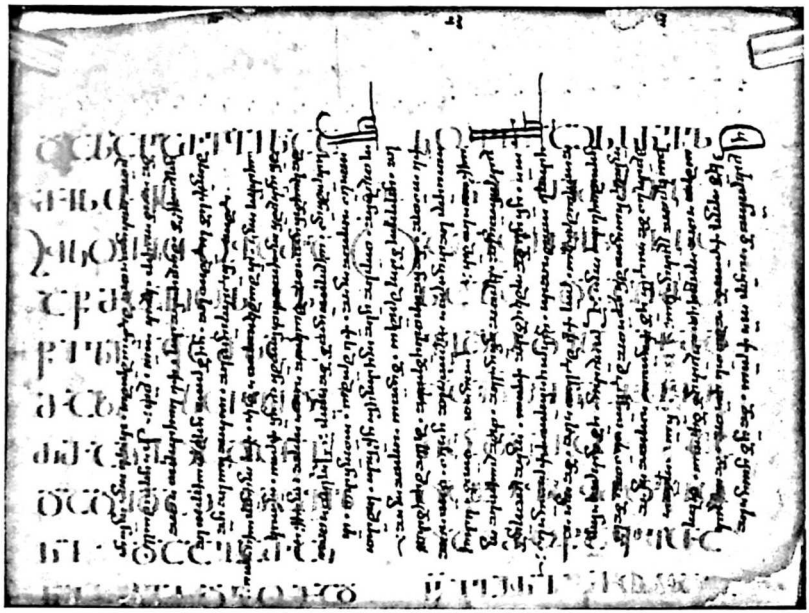
შდრ. ორიგინალური გამოსახულება (ფერადი ფოტო):



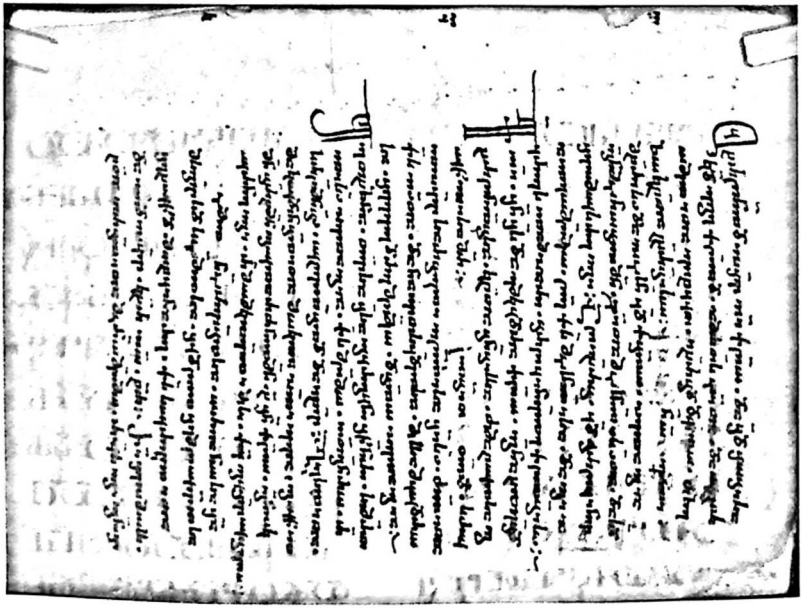
ულტრაიისფერი გამოსახულება:

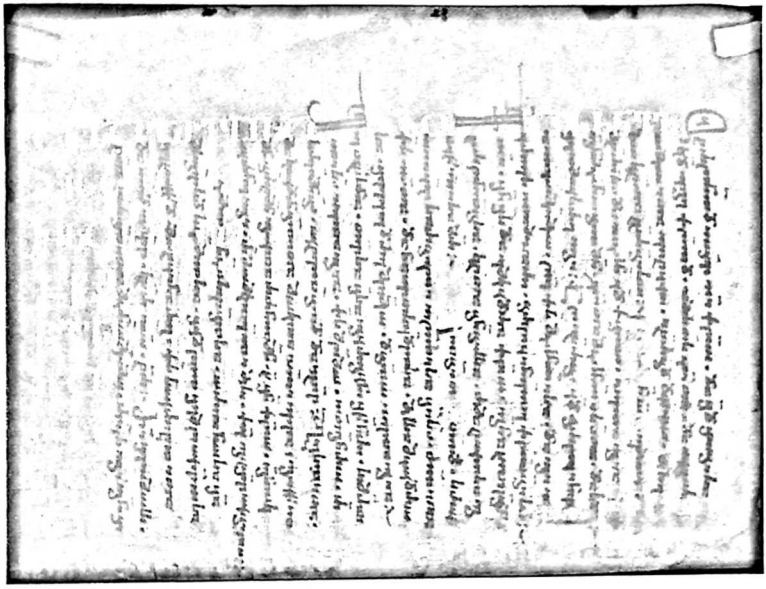


იისფერი გამოსახულება (440nm)



ყვითელი გამოსახულება (580nm)





ზედა გამოსახულების „მოხსნა“ (გაუფერულება)



დიგიტალური ტექნოლოგიების გამოყენებით გაშიფრული ვენის პალიმფსესტი 2007 წელს გამოსცა ბრეპოლსის გამომცემლობამ იოსტ გიპერტის, ზურაბ სარჯველაძისა და ლამარა ქაჯაიას რედაქციით.[148] საგანგებოდ გვინდა აღვნიშნოთ, რომ პალიმფსესტების უახლესი ტექნოლოგიებით დამუშავებამ 98%-მდე გაზარდა ორიგინალური ტექსტის ზედა ფენის გაშიფვრა. ქართულ ხელნაწერთა დიგიტალიზაციაში უდიდესი წვლილი მიუძღვის პროფ. იოსტ გიპერტს, რომლის კვლევის შედეგები დაწვრილებით არის წარმოდგენილი სტატიებში „მრავალსპექტრული გამოსახულების გამოყენება კავკასიური პალიმფსესტის შესწავლისათვის“[149] და „პალიმფსესტების მრავალსპექტრული ანალიზის ახალი პერსპექტივები“.[150]

3. ვირტუალური ბიბლიოთეკის კონცეფცია

ქართული ხელნაწერების უმეტესი ნაწილი საქართველოში ინახება – ძირითადად კორნელი კეკელიძის ხელნაწერთა ეროვნულ ცენტრში, მუზეუმებში, ზოგიერთი კი პირად კოლექციებში. ხელნაწერთა არცთუ მცირე რაოდენობა საქართველოს ფარგლებს გარეთ არის გატანილი. ისტორიული უკუღმართობის გამოისობით ამგვარ ხელნაწერებს, არც თუ იშვიათად, შლიდნენ და ნაწილ-ნაწილებად ყიდდნენ. ასეთი ხელნაწერის ტექსტის მეცნიერული კვლევა დიდ სირთულეებთან იყო დაკავშირებული: საჭირო იყო მათი აღმოჩენა, ფრაგმენტების მისადაგება ძირითად ტექსტთან და ტექსტის „გაერთიანება“. ამგვარად დანაწევრებული ხელნაწერის ტექსტის ანალიზი მკვლევრისაგან მოითხოვდა დიდ დროს და სხვადასხვა ქვეყნებში მოგზაურობას ხელნაწერების ადგილზე კვლევის მიზნით.

საყოველთაო დიგიტალიზაციის პროცესმა, რომელმაც ხელნაწერების დამუშავების სფეროშიც შეაღწია, მნიშვნელოვნად გაუადვილა მეცნიერებს კვლევა: დანაწევრებული ნაწილების დიგიტალიზაციამ საშუალება მისცა მეცნიერებს შეექმნათ ხელნაწერის „ვირტუალური ნუსხა“, რომელიც ორიგინალური ხელნაწერისაგან განსხვავებით ერთ მთლიან ტექსტად არის შეკრული. ამის ერთ-ერთ თვალსაჩინო მაგალითია ერთ-ერთი უძველესი ხელნაწერის ე.წ. Codex Sinaiticus[151] ვირტუალური ნუსხა, რომელიც რუსეთის მეცნიერთა საერთაშორისო თანამშრომლობის შედეგად შეიქმნა.

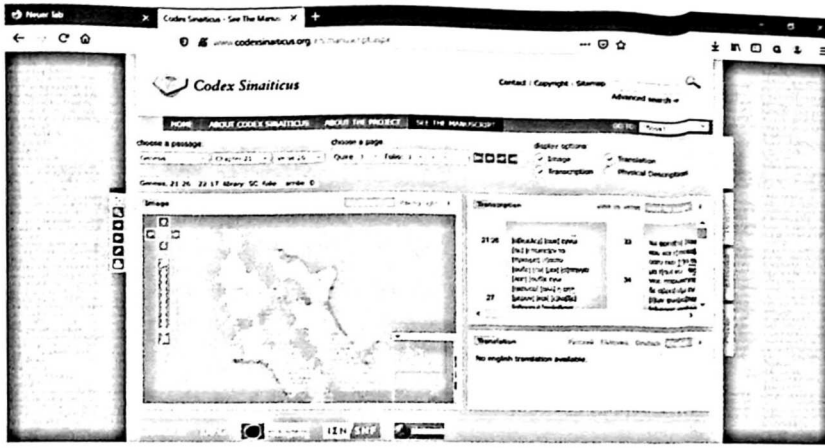
---

[148] იხ. The Old Georgian Palimpsest Codex Vindobonensis georgicus 2, edited by Jost Gippert in co-operation with Zurab Sarjveladze and Lamara Kajaia; Turnhout: Brepols 2007 (Monumenta Palaeographica Medii Aevi, Series Ibero-Caucasica, 1); 368 pp.

[149] http:// titus.fkidg1.uni-frankfurt.de/personal/jg/pdf/jg2007b.pdf

[150] http://titus.fkidg1.uni-frankfurt.de/personal/jg/pdf/jg2013g.pdf

[151] http://www.codexsinaiticus.org/en/



პროექტის უშუალო პარტნიორები იყვნენ:

1. ბრიტანეთის ბიბლიოთეკა
2. ლაიფციგის უნივერსიტეტის ბიბლიოთეკა
3. სინას მთის წმინდა ეკატერინეს მონასტერი
4. რუსეთის ეროვნული ბიბლიოთეკა, სანკტ-პეტერბურგი

პროექტის განხორციელებაში მონაწილეობას იღებდნენ აგრეთვე ბირმინგემის უნივერსიტეტი (ინგლისი),[152] მიუნსტერის უნივერსიტეტი (გერმანია),[153] გოტინგენის უნივერსიტეტი (გერმანია)[154] და ბიბლიური ლიტერატურის საზოგადოება (ატლანტა, ამერიკის შეერთებული შტატები)[155].

პროექტს აფინანსებდნენ შემდეგი ორგანიზაციები:

- The Arts and Humanities Research Council, UK
- Deutsche Forschungsgemeinschaft (German Research Foundation)
- Stavros S. Niarchos Foundation
- Leventis Foundation
- Mariposa Foundation
- J. F. Costopoulos Foundation
- Hellenic Foundation
- American Friends of Saint Catherine's Monastery
- American Trust for the British Library

[152] Institute for Textual Scholarship and Electronic Editing (ITSEE), University of Birmingham, United Kingdom
[153] Institut für neutestamentliche Textforschung, University of Münster, Germany
[154] Göttinger DigitalisierungsZentrum (Centre for Retrospective Digitization), Göttingen State and University Library, Germany
[155] Society of Biblical Literature, Atlanta, USA

ისტორიულ ქარტეხილებს ვერც ქართული ხელნაწერები გადაურჩა. არაერთი ხელნაწერი დაიშალა და დანაწევრებული სახით გაიფანტა სხვადასხვა კონტინენტზე. ამის ერთ-ერთი თვალსაჩინო მაგალითია საბაწმინდაზე გადაწერილი ხელნაწერი – საბაწმინდის ლექციონარი. აღნიშნული ლექციონარი ორი თვალსაზრისით არის მნიშვნელოვანი:

1. ლექციონარი განეკუთვნება ჩვენამდე მოღწეულ იმ მცირე ხელნაწერთა რიცხვს, რომელიც ხანმეტ ფორმებს შეიცავს;
2. ქართულად თარგმნილი ეს ლექციონარი, რომელსაც "იერუსალიმურ კანონსაც" უწოდებენ, უძველესია და ასახავს V-VIII სს-ის იერუსალიმის ღვთისმსახურების წესს. ქართულად თარგმნილი ლექციონარის ბერძნული დედანი დაკარგულია, ამდენად ამ სახის ქართულ კრებულებს განსაკუთრებული მნიშვნელობა ენიჭება ადრეული ბერძნული, ანუ იერუსალიმური ღვთისმსახურების შესწავლაში.

ტრაგიკულია საბაწმინდის ლექციონარის ბედი. ჯერ კიდევ XIX ს-ის დასასრულს იგი სინის მთის წმ. ეკატერინეს მონასტრის ქართულ კოლექციაში ინახებოდა. 29 ფურცლისაგან შემდგარი ფრაგმენტი 1883 წელს აღწერა სინის მთაზე მყოფმა ალ. ცაგარელმა. XX ს-ში წიგნი დაშლილი სახით მოხვდა ევროპის წიგნსაცავებში: 27 ფურცელი – ავსტრიაში, გრაცის უნივერსიტეტის წიგნსაცავში, 1 ფურცელი – ბრიტანეთში, ბირმინგემის Sellyak College-ის ბიბლიოთეკაში და 1 ფურცელი – საფრანგეთის ნაციონალურ ბიბლიოთეკაში, პარიზში. საინტერესოა, რომ VII ს-ით დათარიღებული ეს ფურცლები საკმაოდ კარგადაა მოღწეული. ისინი არ გადაუშლიათ და არ გამოუყენებიათ ხელმეორე საწერ მასალად, ამოფხეკილია მხოლოდ ხმარებიდან გამოსული "ხანები". როგორც ჩანს, იერუსალიმის წიაღიდან გამოსულმა ამ წიგნმა არც შემდგომ განიცადა რეციზია და კვლავ აქტიური გამოყენების კრებულს წარმოადგენდა.

ხელნაწერის დიგიტალური ვერსიის შექმნამ შესაძლებელი გახადა ხელნაწერის ტექსტის „ვირტუალური გამთლიანება".

ხელნაწერების ვირტუალურად გამთლიანებამ, ისევე როგორც ვირტუალური ბიბლიოთეკების შექმნამ ბევრად გააადვილა საკვლევ რესურსებზე წვდომა და დააჩქარა კვლევის პროცესი ტექნიკური თვალსაზრისით.

დანართი II

კაუკასიოლოგია მაინის ფრანკფურტის უნივერსიტეტში

I. ფრანკფურტის უნივერსიტეტის კაუკასიოლოგიის სასწავლო პროგრამა

ქართველური და კაუკასიური ენები 2002 წლიდან ისწავლება ფრანკფურტის გოეთეს სახ. უნივერსიტეტში შედარებითი ენათმეცნიერების ინსტიტუტში. 2005 წლიდან იგი ჩამოყალიბდა კაუკასიური ენათმეცნიერების განყოფილებად ემპირიული ენათმეცნიერების ინსტიტუტის ფარგლებში და ამ სახით განაგრძობს არსებობას დღემდე. 2007 წლიდან აკრედიტაცია გაიარა კაუკასიური ენათმეცნიერების ბეჭლერის პროგრამამ, 2011 წლიდან კი სამაგისტრო პროგრამამ კაუკასიურ ენებში. 2013 წელს განხორციელდა სამიეე აკადემიური საფეხურის – ბაკალავრიატის, მაგისტრანტურისა და დოქტორანტურის ხელახალი აკრედიტაცია.

სასწავლო პროგრამა მოიცავს შემდეგ სალექციო კურსებსა და სემინარებს:

1. კაუკასიური ენათმეცნიერების შესავალი;
2. კაუკასიის ხალხები და ენები;
3. შედარებითი ქართველოლოგია;
4. ქართველური ენები: ქართული, მეგრული, ლაზური, სვანური;
5. კაუკასიური ენები: ხუნძური, უდიური, აფხაზური და კაუკასიური ალბანური;
6. კაუკასიოლოგიის სპეციალური პრობლემები: ერგატიულობა კაუკასიურ ენებში, უარყოფის კატეგორია ქართულში, რეფერენციულობა ქართულში, ევიდენციალობა, მოდალობის კატეგორია ქართულში, სახელური მოდიფიკაციები ქართულში, ქართული ზმნის სემანტიკური ანალიზი;
7. ენობრივი კონტაქტები და ბილინგვიზმი;
8. საფრთხეში მყოფი კაუკასიური ენების ელექტრონული დოკუმენტირება;
9. კაუკასიურ ენათა დიგიტალური ბანკები;
10. დიგიტალური ქართველოლოგია;
11. დიგიტალური რუსთველოლოგია.

კაუკასიურ ენათა სწავლება მიმდინარეობს გერმანულ უმაღლეს საგანმანათლებლო სისტემაში დანერგილი ინოვაციური მეთოდების გამოყენებით – სასწავლო კონცეფციის საფუძველზე „24 საათიანი სასწავლო სიერცე". 2011 წლიდან სასწავლო პროცესში გამოყენებულია ელექტრონული სწავლების პლატფორმა OLAT (Online Learning And Training), ხოლო 2013 წლიდან eLecture (ელექტრონული ლექციები).

II. კავკასიოლოგიაში დაცული დისერტაციები ფრანკფურტის უნივერსიტეტში

ფრანკფურტის უნივერსიტეტში მომზადდა 7 სადოქტორო დისერტაცია ქართველოლოგიასა და კავკასიოლოგიაში:

1. ირინა ცინცაძე „Die Entwicklungstendenzen des Passivs in der georgischen Schriftsprache", 2005
2. სვეტლანა ალბორნი, "Verfahren zur Markierung von Fragesätzen", 2008
3. ზაქარია ფურცხვანიძე, "Fokuspartikeln und Wortstellung im Georgischen", 2011
4. ალექსანდრე როსტოვცევ-პოპიელი, "Grammaticalized Affirmativity in Kartvelian", 2012
5. ლელა სამუშია, "Hypotaxe in den Kartwelsprachen", 2014
6. ნათია დუნდუა, „ტობის წიგნის" ქართული თარგმანის წარმომაელობა, 2017
7. მარიამ ყამარაული, „სახელური მოდიფიკაციები ქართულში", 2018

ფრანკფურტის უნივერსიტეტის კავკასიოლოგიის ფარგლებში დაარსდა სამეცნიერო ჟურნალი **„Folia Caucasica"** (გამომცემლები პროფ. იოსტ გიპერტი და პროფ. მანანა თანდაშვილი), სადაც ძირითადად იბეჭდება ფრანკფურტის უნივერსიტეტში დაცული დისერტაციები კავკასიოლოგიაში. ჟურნალი **„Folia Caucasica"** გამოიცემა გერმანულ გამომცემლობაში Reichert-Verlag. დღემდე გამოცემულია ოთხი ტომი:

1. 2011: **Folia Caucasica I**: Festschrift zum 55. Geburtstag von Prof. J. Gippert, Frankfurt/Tbilissi, Logopress (herausgegeben von Manana Tandaschwili & Zakharia Pourtskhvanidze)
2. 2014: **Folia Caucasica II**: Zakharia Pourtskhvanidze, Fokuspartikeln und Wortstellung im Georgischen, Reichert Verlag, Wiesbaden (herausgegeben von Jost Gippert & Manana Tandaschwili)
3. 2018: **Folia Caucasica III:** Irine Zinzadze, Die Entwicklungstendenzen des Passivs in der georgischen Schriftsprache, Reichert Verlag, Wiesbaden (herausgegeben von Jost Gippert & Manana Tandaschwili)
4. 2018: **Folia Caucasica IV**: Lela Samuschia, Hypotaxe in den Kartwelsprachen, Reichert Verlag, Wiesbaden (herausgegeben von Jost Gippert & Manana Tandaschwili)

ამჟამად გამოსაცემად მზადდება **Folia Caucasica**-ს **V** და **VI** ტომები – ნათია დუნდუასა და მარიამ ყამარაულის სადოქტორო დისერტაციები.

III. ფრანკფურტის უნივერსიტეტში განხორციელებული საერთაშორისო პროექტები კავკასიოლოგიის დარგში

დაფინანსება: გერმანული (Volkswagen-Stiftung), ევროპული (INTAS, TEMPUS) და ქართული (შოთა რუსთაველის სახ. ეროვნული სამეცნიერო ფონდი) ფონდების მიერ.

1. 1995-1997: „Georgisches Verbum" (in Zusammenarbeit mit D. Melikishvili, Tbilisi, und G. Sharashidze, Paris; INTAS, ca. EUR 50.000,-)
2. 1999-2002: ARMAZI-Projekt („Kaukasische Sprachen und Kulturen: Elektronische Dokumentation"; Volkswagen-Stiftung, ca. EUR 117.900,-);
3. 2002-2005: ECLinG-Projekt („Bedrohte kaukasische Sprachen in Georgien"; Volkswagen-Stiftung, EUR 167.800,-);
4. 2004-2006: Kaukasische Palimpseste („Neue Wege zur wissenschaftlichen Bearbeitung von Palimpsesthandschriften kaukasischer Provenienz" (in Zusammenarbeit mit W. Schulze, München / Z. Aleksidze, Tbilisi / J.-P- Mahé, Paris; Volkswagen-Stiftung, 155.000,- EUR);
5. 2005-2007: SSGG („Die sprachliche Situation im gegenwärtigen Georgien"; Volkswagen-Stiftung, EUR 189.000,-);
6. 2006-2009: Old Georgian Gospels ("Critical Edition of the Old Georgian Version of Matthew's and Mark's Gospels – Catalogue of the Manuscripts Containing the Old Georgian Translation of the Gospels"; in Zusammenarbeit mit D. Tvaltvadze, Tbilisi / B. Outtier, Paris; INTAS, EUR 110.000,-);
7. 2009-2012: Georgische Palimpsesthandschriften (in Zusammenarbeit mit D. Tvaltvadze, Tbilisi/ B. Outtier, Paris; Volkswagen-Stiftung, EUR 99.000,-);
8. 2011-2012: Corpus Caucasicum (Google, ca. EUR 37.000,-)
9. 2011-2014: Kaukasische Evangelien / Die Überlieferung der vier Evangelien in altgeorgischer Sprache / Kaukasische Palimpseste (innerhalb des LOEWE-Schwerpunkts „Digital Humanities"; ca. EUR 250.000,-)
10. 2011-2015: Documentation of Khinalug (in Zusammenarbeit mit W. Schulze, München; Volkswagen-Stiftung, ca. EUR 350.000,-);
11. 2012-2015: GNC („Georgisches Nationalcorpus"; Volkswagen-Stiftung, EUR 390.400,-);
12. 2012-2015: დრო და სივრცე ქართველურ ენებში (პროფ. ბელა ქობალავასთან თანამშრომლობით, თბილისი, რუსთაველის ეროვნული სამეცნიერო ფონდი, 149.941,- ლარი).
13. 2016-2019: ქართული ენის ეროვნული კორპუსი GNC – თემატური კორპუსები („Georgisches Nationalcorpus"; Volkswagen-Stiftung, EUR 60.000,-)

14. 2016-2019: ლინგვოკულტუროლოგიური დიგიტალური არქივი (პროფ. რამაზ ხალვაშთან თანამშრომლობით, თბილისი, რუსთაველის ეროვნული სამეცნიერო ფონდი, 290.000,- ლარი).

15. 2016-2017: საერთაშორისო სადოქტორო სკოლის მოსამზადებელი პროგრამა დიგიტალურ ჰუმანიტარიაში (Volkswagen-Stiftung, EUR 96.000,)

განხორციელებული პროექტების საერთო ღირებულება დაახლოებით **2** მილიონი ევროა. საქართველოდან პროექტში ჩართული იყო 17 სამეცნიერო-კვლევითი ინსტიტუტი, უნივერსიტეტი და მუზეუმი. აღნიშნული სამეცნიერო პროექტების განხორციელებაში მონაწილეობას იღებდა 500-მდე პროფესორი, დოცენტი, დოქტორანტი და სტუდენტი, ბიბლიოთეკისა და მუზეუმის თანამშრომლები.

IV. საერთაშორისო კონფერენციები კავკასიოლოგიაში:

საერთაშორისო სამეცნიერო ფონდის Volkswagen Stiftung-ის (გერმანია) დაფინანსებით ფრანკფურტის უნივერსიტეტის კავკასიოლოგიის ორგანიზებით სამჯერ ჩატარდა საერთაშორისო სამეცნიერო კონფერენცია ფრანკფურტის უნივერსიტეტში:

- 12-14.10. 2003: Internationale Konferenz „Die sprachliche Situation im gegenwärtigen Georgien", Frankfurt a/M (EUR 10.000,-)

- 5.-10.9.2004: Internationale Konferenz „Dokumentation bedrohter Sprachen", Frankfurt a/M (EUR 35.000,-)

- 17-21.05. 2010: Internationale Konferenz "Zwischen Europa und Orient – Mittelasien/Kaukasus im Fokus der Wissenschaft, Tbilisi (EUR 196.000,-);

პროექტების საერთო ღირებულება შეადგენდა 241 000 ევროს.

V. ფრანკფურტის ლინგვისტური წრე და მის მიერ საქართველოში ორგანიზებული საზაფხულო სკოლები დიგიტალურ ჰუმანიტარიაში

2010 წლიდან ფრანკფურტის უნივერსიტეტის ემპირიული ენათმეცნიერების ინსტიტუტში ფუნქციონირებს **ფრანკფურტის ლინგვისტური წრე** (დამაარსებელი და ხელმძღვანელი პროფ. მანანა თანდაშვილი), რომელიც საშუალებას აძლევს კავკასიოლოგიის სტუდენტებს სწავლების ჯერ კიდევ ადრეულ ფაზაში აქტიურად ჩაერთონ სამეცნიერო-კვლევით სამუშაოებში.

2012 წლიდან ფრანკფურტის ლინგვისტური წრის ინიციატივით სისტემატურად ტარდება საქართველოში საზაფხულო სკოლები დიგიტალური ჰუმანიტარიის სფეროში:

- 2012: ბათუმის I საზაფხულო სკოლა: „დიგიტალური ჰუმანიტარია – ენათა ელექტრონული დოკუმენტაცია და ქართველოლოგია“ (პროფ. რ. ხალვაშთან ერთად)

- 2013: ბათუმის II საზაფხულო სკოლა თემაზე: „დიგიტალური ჰუმანიტარია – კორპუსლინგვისტიკა და ქართველოლოგიის სპეციალური პრობლემები“ (პროფ. რ. ხალვაშთან ერთად)

- 2014: ბაკურიანის ზამთრის სკოლა თემაზე: „დიგიტალური ჰუმანიტარია – ქართველოლოგია და 21 საუკუნის გამოწვევები“ (პროფ. დ. თვალთვაძესთან ერთად)

- 2014: ბათუმის III საზაფხულო სკოლა თემაზე: „დიგიტალური ჰუმანიტარია – კორპუსლინგვისტიკა და ქართველოლოგიის სპეციალური პრობლემები“ (პროფ. რ. ხალვაშთან ერთად)

- 2015: ბათუმის IV საზაფხულო სკოლა თემაზე: „დიგიტალური ჰუმანიტარია – ენის დოკუმენტაცია“ (პროფ. რ. ხალვაშთან ერთად)

- 2015: ვორკშოფი დიგიტალურ ჰუმანიტარიაში თემაზე: „კორპუსის აგების ძირითადი პრინციპები და მეთოდები“ (პროფ. ი. გიპერტთან ერთად)

- 2016: თბილისის ზამთრის სკოლა თემაზე: „დიგიტალური ჰუმანიტარია – თემატური კორპუსების აგების ძირითადი პრინციპები“ (პროფ. დარეჯან თვალთვაძესთან ერთად)

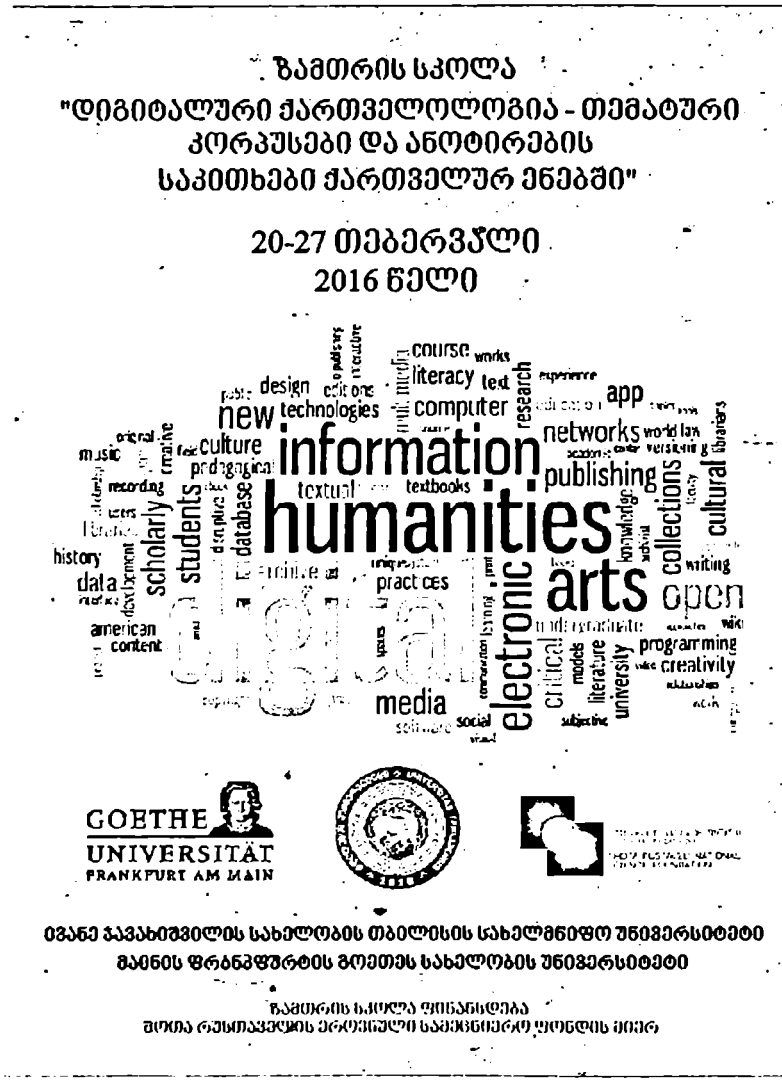
- 2016: ბათუმის V საზაფხულო სკოლა თემაზე: „დიგიტალური ჰუმანიტარია – ენისა და კულტურის დოკუმენტირება“ (პროფ. რ. ხალვაშთან ერთად)

- 2017: ბათუმის VI საზაფხულო სკოლა თემაზე: „დიგიტალური ჰუმანიტარია – ენისა და კულტურის დოკუმენტირება“ (პროფ. რ. ხალვაშთან ერთად)

- 2018: ბათუმის VII საზაფხულო სკოლა თემაზე: „დიგიტალური ჰუმანიტარია – ენისა და კულტურის დოკუმენტირება“ (პროფ. რ. ხალვაშთან ერთად)

- 2019: ბათუმის საერთაშორისო VIII სასაფხულო სკოლა დიგიტალურ კაეკასიოლოგიაში – პირველი ეტაპი (პროფ. ი. გიპერტთან, პროფ. დ. ფორკერთან, და პროფ. რ. ხალვაშთან ერთად)

- 2020: ბათუმის IX სასაფხულო სკოლა დიგიტალურ ქართველოლოგიაში (პროფ. რ. ხალვაშთან ერთად)





გარდა ქართველი მეცნიერებისა, საზაფხულო სკოლების პროგრამაში მონაწილეობას იღებდნენ უცხოელი მეცნიერები: იოსტ გიპერტი (ფრანკფურტი), ირინა გურევიჩი (დარმშტადტი), ნიკოლაუს პიმელმანი (კიოლნი), ნოა ბუბენჰოფერი (ციურიხი), გერჰარდ პაიერი (ლაიფციგი), დიანა ფორკერი (იენა), მიშა დანიელი (მოსკოვი), ნინა დობრუშინა (მოსკოვი), სვენ გრავუნდა (ფრანკფურტი), ქრისტოფერ დრაქსლერი (მიუნხენი), უვე კვასტჰოფი (ლაიფციგი), ქრისტიან ჩიარკოსი (ფრანკფურტი) და სხვ.

საზაფხულო სკოლებს აფინანსებდნენ: Volkswagen Stiftung (გერმანია), შოთა რუსთაველის სახ. ეროვნული სამეცნიერო ფონდი, საქართველოს მეცნიერებისა და განათლების სამინისტრო, აჭარის მეცნიერების, განათლებისა და სპორტის სამინისტრო. დაფინანსება მთლიანობაში შეადგენდა დაახლოებით 120 000 ევროს.

ფრანკფურტის ლინგვისტურ წრეს უდიდესი წვლილი მიუძღვის საქართველოში დიგიტალური ჰუმანიტარიის გავრცელებისა და დანერგვის თვალსაზრისით. გარდა საზაფხულო სკოლებისა 2015 წელს, პირველად საქართველოში, ქართულ ენაზე მომზადდა და აკრედიტაცია გაიარა ბათუმის უნივერსიტეტში **მაინორმა პროგრამამ დიგიტალურ ჰუმანიტარიაში.** პროგრამას ხელმძღვანელობს ფრანკფურტის უნივერსიტეტის პროფესორი კავკასიოლოგიის დარგში მანანა თანდაშვილი.

ფრანკფურტის უნივერსიტეტის კავკასიოლოგია აქტიურად იღებდა მონაწილეობას ქართველოლოგიის ინტერნაციონალიზაციის პროცესში, კერძოდ სამეცნიერო-აკადემიური კადრების ურთიერთგაცვლის საერთაშორისო პროგრამებში. გერმა-

ნული და ქართული სამეცნიერო ფონდების დაფინანსებით ფრანკფურტის უნივერსიტეტში სამეცნიერო მიელინებით იმყოფებოდნენ ქართველი მეცნიერები და დოქტორანტები ქართველოლოგიისა და კავკასიოლოგიის დარგში:

1995-96: პროფ. ვახტანგ იმნაიშვილი (ივ. ჯავახიშვილის თბილისის სახელმწიფო უნივერსიტეტი; DFG, ორსემესტრიანი მოწვევა)

1996-97: პროფ. ზურაბ სარჯველაძე (ივ. ჯავახიშვილის თბილისის სახელმწიფო უნივერსიტეტი; DFG, ერთსემესტრიანი მოწვევა)

2000: პროფ. თამაზ გამყრელიძე (საქართველოს მეცნიერებათა აკადემია; Volkswagen-Stiftung, ერთსემესტრიანი მოწვევა)

2005-07: დოქტ. ელგუჯა დადუნაშვილი (ლიტერატურის ინსტიტუტი, Tbilisi; Volkswagen-Stiftung, ორწლიანი მოწვევა)

2009-10: პროფ. მიხეილ ქურდიანი (არნ. ჩიქობავას სახ. ენათმეცნიერების ინსტიტუტი: DAAD, Projekt: „ბასკურ-კავკასიური ურთიერთობები", სამთვიანი მოწვევა)

2012: პროფ. ნინო შარაშენიძე (ივ. ჯავახიშვილის თბილისის სახელმწიფო უნივერსიტეტი; DAAD, Projekt: „მოდალობა ქართულში", ორთვიანი მოწვევა)

2012: პროფ. მაია ლომია (ილიას უნივერსიტეტი, შოთა რუსთაველის ეროვნული სამეცნიერო ფონდი: „მეგრულის ლინგვისტური ანალიზი", სამეცნიერო მიელინება)

2012: პროფ. რუსუდან გერსამია (ილიას უნივერსიტეტი, შოთა რუსთაველის ეროვნული სამეცნიერო ფონდი: „მეგრულის ლინგვისტური ანალიზი", სამეცნიერო მიელინება)

2013: პროფ. მაია ლომია (ილიას უნივერსიტეტი, შოთა რუსთაველის ეროვნული სამეცნიერო ფონდი: „დრო და სივრცე ქართულში, სამეცნიერო მიელინება)

2013: პროფ. რუსუდან გერსამია (ილიას უნივერსიტეტი, შოთა რუსთაველის ეროვნული სამეცნიერო ფონდი: „დრო და სივრცე ქართულში", სამეცნიერო მიელინება)

2014: პროფ. გიული შაბაშვილი, (ივ. ჯავახიშვილის თბილისის სახელმწიფო უნივერსიტეტი; შოთა რუსთაველის ეროვნული სამეცნიერო ფონდი: "პერფექტულობა ქართულში", სამთვიანი მოწვევა)

2015: პროფ. მაია ლომია (ილიას უნივერსიტეტი, შოთა რუსთაველის ეროვნული სამეცნიერო ფონდი: „დრო და სივრცე ქართულში", სამეცნიერო მიელინება)

2015: პროფ. რუსუდან გერსამია (ილიას უნივერსიტეტი, შოთა რუსთაველის ეროვნული სამეცნიერო ფონდი: „დრო და სივრცე ქართულში“, სამეცნიერო მიმართულება)

2015: დოქტორანტი ნათია ფურცელაძე (“გრიგოლ რობაქიძის არქივი“, შოთა რუსთაველის ეროვნული სამეცნიერო ფონდი, სამთვიანი მივლინება).

2018: დოქტორანტები დიანა კაკაშვილი და დავით თინიკაშვილი (DAAD, სამთვიანი მივლინება)

გარდა ბათუმის შოთა რუსთაველის სახელმწიფო უნივერსიტეტისა, ფრანკფურტის უნივერსიტეტის კავკასიოლოგია აქტიურად თანამშრომლობს ივ. ჯავახიშვილის სახ. თბილისის სახელმწიფო უნივერსიტეტთან. 2018 წლის 11 ოქტომბერს გაფორმდა ურთიერთთანამშრომლობის მემორანდუმი ფრანკფურტის უნივერსიტეტსა და თბილისის სახელმწიფო უნივერსიტეტს შორის, რომელიც ითვალისწინებდა თანამშრომლობას ერთობლივი საერთაშორისო სამეცნიერო პროექტების განხორციელების, საასპირანტო სკოლების ორგანიზებისა და დოქტორანტთა გაცვლის თვალსაზრისით. აღნიშნული მემორანდუმის ფარგლებში 2019-2020 წლებში თსუ-ში ერთწლიანი სტიპენდიით (DAAD) იმყოფებოდა ფრანკფურტის უნივერსიტეტის კავკასიოლოგიის დოქტორანტი ნიდა ონაკი.

V. ქართული ლიტერატურა ფრანკფურტში

გარდა სამეცნიერო აკადემიური აქტივობებისა, ფრანკფურტის უნივერსიტეტის კავკასიოლოგია წლების განმავლობაში უწყობდა ხელს ქართული ლიტერატურის პოპულარიზაციას გერმანულენოვან ლიტერატურულ სივრცეში. ჯერ კიდევ 2008 წელს დააარსა პროფ. მანანა თანდაშვილმა **“ქართული ლიტერატურის სალონი – ევტერპე“**, რომელმაც 10 წელი იარსება ფრანკფურტში და 140-ზე მეტი ლიტერატურული საღამო ჩაატარა ფრანკფურტის რომანის ფაბრიკაში, წიგნის მაღაზიებსა, ლაიფციგისა და ფრანკფურტის წიგნის ბაზრობაზე. 2015 წელს მანანა თანდაშვილის ინიციატივით ფრანკფურტის უნივერსიტეტში გაიხსნა შოთა რუსთაველის სახელობის ქართული ლიტერატურის კაბინეტი, რომელიც დღეისათვის შოთა რუსთაველის „ვეფხისტყაოსნის“ თარგმანების უდიდეს კოლექციას ფლობს სხვადასხვა ენაზე.

ფრანკფურტის ლიტერატურულ სალონთან თანამშრომლობით, ფრანკფურტის უნივერსიტეტის კავკასიოლოგიამ 2018 წელს აქტიური მონაწილეობა მიიღო ფრანკფურტის წიგნის ბაზრობაზე ქართული პროგრამის მომზადებაში: ფრანკფურტის წიგნის ბაზრობაზე გერმანიაში გამოცემული 19 წიგნი წარადგინა:

1. **Georgische Gegenwartsliteratur**, Reichert Verlag, Wiesbaden, 2010
2. **Techno der Jaguare** (Neue Erzählerinnen aus Georgien), Frankfurter Verlagsanstalt, Frankfurt a/M, 2012
3. **Beso Chwedelidze**, Der Geschmack von Asche (Erzählband), Literaturverlag Leipzig, 2013
4. **Dato Turaschwili**, Westflug, Wagenbach Verlag, Berlin, 2014
5. **Schota Rustaweli**, Der Recke im Tigerfell, Reichert Berlag, Wiesbaden, 2014
6. **Zaza Burchuladze**, Adibas, Aufbau Verlag, Berlin, 2015
7. **Guram Odisharia**, Der Pass der Flüchtlinge, Reichert Verlag, 2015
8. **Georgische Dramaturgie**, Theater der Zeit, Berlin, 2015
9. **Paata Berikaschwili**, Der Engel genannt Sali, Frankfurter Verlagsgruppe, 2015
10. **Bondo Matsaberidze**, Das Märchen von Bekna und Thekla, Drava Verlag, 2017
11. **Bondo Matsaberidze**, Das Märchen von Bussi und Thussi, Drava Verlag, 2018
12. **Goderdzi Tschocheli**, Der scharlachrote Wolf,), Frankfurter Verlagsanstalt, 2018
13. **Khatuna Tavdgiridze**, Der Fisch mit zwei Schatten, Wieser Verlag, Wien, 2018
14. **Frauenstimmen aus Georgien**, Pop Verlag, Ludwigsburg, 2018
15. **Liederbuch Georgien**, Reichert Berlag, Wiesbaden, 2018
16. **Georgische Handschriften**, Reichert Berlag, Wiesbaden, 2018
17. **Guram Odischaria, Georgien, Tbilisi, Rustaweli-Boulevard**, Reichert Berlag, Wiesbaden, 2018
18. **Ilia Tschawtschawadze**, Erzählingen aus Georgien, Reichert Berlag, Wiesbaden, 2018
19. **Gutcha Kvarakhelia, Die Lichtszrahlen**, Bakmi, 2018

გამოყენებული ლიტერატურა:

1. ხათუნა ბერიძე, გრიგოლ კახიანი, პოლიტიკური ტექსტების ქართული მთარგმნელობითი კორპუსი, საერთაშორისო კონფერენცია „ქართული ენა და თანამედროვე ტექნოლოგიები“, თეზისები, თბილისი, 2013, გვ. 25-27.
2. იოსტ გიპერტი, სიახლენი ხანმეტობიდან, ენათმეცნიერების საკითხები I-II, 2009, გვ. 164-184.
3. ნინო დობორჯგინიძე, ირინა ლობჟანიძე, ქართლის ცხოვრების პარალელური კორპუსი, მიმოხილვა. file:///C:/Users/Manana/AppData/Local/Temp/233-Article%20Text-467-1-10-20170101.pdf
4. ივანე იმნაიშვილი, ქართული ოთხთაეის სიმფონის-ლექსიკონი, რედ. აკ. შანიძე, თბილისი: საქართველოს სსრ მეცნიერებათა აკადემიის გამომცემლობა, 1948-1949.
5. მანანა თანდაშვილი, ენისა და კულტურის დოკუმენტირების თანამედროვე მეთოდები და ინსტრუმენტები“, თბილისი, გამომცემლობა უნივერსალი, 2020 წ.;
6. მანანა თანდაშვილი, რამაზ ხალვაში, რუსუდან პაპიაშვილი, მზია ხახუტაიშვილი, გრიგოლ კახიანი, ნანა ცეცხლაძე, ენისა და კულტურის ელექტრონული დოკუმენტირების თანამედროვე ინსტრუმენტები (მეგზური), თბილისი, გამომცემლობა უნივერსალი, 2020 წ.
7. მანანა თანდაშვილი, ზაქარია ფურცხვანიძე, კორპუსლინგვისტიკის გლოსარიუმი, ფრანკფურტი-თბილისი, 2014.
8. მანანა თანდაშვილი, ზაქარია ფურცხვანიძე, „ქართული ენის კორპუსლინგვისტური პარადიგმა“, ფრანკფურტი-თბილისი, ინტელექტი, 2014.
9. თინათინ მარგალიტაძე, ევროპულ-ქართული პარალელური კორპუსები ქართული ლექსიკოგრაფიისა და მთარგმნელობითი საქმიანობისათვის. https://margaliti.com/paralelkorpus.pdf
10. თინათინ მარგალიტაძე, „ინგლისურ-ქართული ლექსიკოგრაფია“, სამეცნიერო ჟურნალი „ქართველოლოგი“, თბილისის სახელმწიფო უნივერსიტეტი, 2013.
11. თინათინ მარგალიტაძე, გ. ქურეჭაშვილი. „ვორდის ფაილებიდან თანამედროვე ონლაინ-ლექსიკონამდე“. საერთაშორისო კონფერენციის კრებული “ელექტრონული ლექსიკოგრაფია 21-ე საუკუნეში (Electronic lexicography in the 21st century: thinking outside the paper)”. ტალინი, 2013. (http://eki.ee/elex2013)
12. თინათინ მარგალიტაძე. „ქართული ტერმინოლოგიური ონლაინლექსიკონების ახალი პლატფორმა და მრავალენოვანი ლექსიკონის მართვის სისტემა“. ევროპის ლექსიკოგრაფთა ასოციაციის (EURALEX) მე-18 საერთაშორისო კონგრესის კრებული. ლიუბლიანას სახელმწიფო უნივერსიტეტის გამომცემლობა. ლიუბლიანა, სლოვენია, 2018. https://euralex.org/wp-content/themes/euralex
13. შოთა რუსთაველი, ვეფხისტყაოსანი, რედ.: აკ. შანიძე, 1975 წ.
14. შოთა რუსთაველი, ვეფხისტყაოსანი, რედ.: ალ. ბარამიძე კ. კეკელიძე, აკ. შანიძე, 1957 წ.

15. ვეფხის-ტყაოსნის სიმფონია. თბილისი: სტალინის სახელობის თბილისის სახელმწიფო უნივერსიტეტის გამომცემლობა, 1956.
16. თედო უთურგაიძე, გრამატიკული კატეგორიებისა და მათი ურთიერთმიმართებისათვის ქართულ ზმნაში, თბილისი, 2002.
17. რამაზ ხალვაში, დიგიტალური ჰუმანიტარიის შესავალი, ბათუმი, 2018.
18. Gabriel Altmann, *Wiederholungen in Texten*. Brockmeyer, Bochum, 1988.
19. William O. Aydelotte, *Quantifizierung in der Geschichtswissenschaft*, in: Hans-Ulrich Wehler (Hrsg.), Geschichte und Soziologie, 2. Aufl. Königstein 1984, S. 259-282.
20. Philipp Balsiger, *Transdisziplinarität: Systematisch-vergleichende Untersuchung disziplinenübergreifen-der Wissenschaftspraxis*, München 2005.
21. Frank Brand, Franz Schaller und Harald Völker (Hg.), *Transdisziplinarität. Bestandsaufnahme und Perspektiven*, Beiträge zur THESIS-Arbeitstagung im Oktober 2003 in Göttingen, Universitätsverlag Göttingen, 2004.
22. Noah Bubenhofer, *Sprachgebrauchsmuster. Korpuslinguistik als Methode der Diskurs- und Kulturanalyse*. Berlin, New York: de Gruyter, 2009.
23. Hadumod Bussmann, *Lexikon der Sprachwissenschaft*, 2008, Kröner, S. 143.
24. Ellen Collins, Michael Jubb, *How do Researchers in the Humanities Use Information Resources?*, in: Liber Quarterly 21 (2012), H. 2, S. 176-187.
25. Klaus Gantert, *H-Soz-u-Kult – Informationsdienst für die Geschichtswissenschaften*, in: Geschichte in Wissenschaft und Unterricht 62 (2011), H. 11/12, S. 645-650.
26. Jost Gippert in co-operation with Zurab Sarjveladze and Lamara Kajaia (eds), *The Old Georgian Palimpsest* Codex Vindobonensis georgicus 2, Turnhout: Brepols 2007 (Monumenta Palaeographica Medii Aevi, Series Ibero-Caucasica, 1).
27. Jost Gippert, *Nochmals zu Beuinschrift von Bolnisi*, in: Kaukasiologie heute, Festschrift für Heinz Fähnrich zum 70. Geburtstag, Jena, Buchverlag König, 2015, 155-182. https://titus.uni-frankfurt.de/personal/jg/pdf/jg2011k.pdf
28. Peter Haber, Jan Hodel, *Geschichtswissenschaft und Web 2.0.* Eine Dokumentation (= The hist.net Working Paper Series; 2), Basel 2010. http://www.histnet.ch/repository/hnwps/hnwps-02.pdf
29. Gertrude Hirsch Hadorn, Holger Hoffmann-Riem, Susette Biber-Klemm, Dominique Joye, Christian Pohl, Urs Wiesmannn, Elisabeth Zemp, (Hg.): *Handbook of Transdisciplinary Research*, Dordrecht 2008.
30. Rüdiger Hohls, Peter Helmberger (Hrsg.), *Humanities-Net Sozial- und Kulturgeschichte* (H-Soz-u-Kult). Bilanz nach 3 Jahren (= Historical Social Research; 24/3), Köln 1999.
31. Dell H. Hymes (Hrsg.), *The Use of Computers in Anthropology* (= Studies in General Anthropology; 2), London 1965.
32. Konrad H. Jarausch, *Möglichkeiten und Probleme der Quantifizierung in der Geschichtswissenschaft*, in: ders. (Hrsg.), Quantifizierung in der Geschichtswissenschaft. Probleme und Möglichkeiten, Düsseldorf 1976, S. 11-30, hier S. 13.
33. Richard Jensen, *Internet's Republic of Letters*: H-Net for Scholars, 1997. http://web.archive.org/web/20000903084512/http://members.aol.com/dann01/whatis.html

34. 1st World Congress of Trandisciplinarity, Preamble. Convento da Arrábida, Portugal, 1994, November 2-6.

35. Mariam Kamarauli, *The Nominal Domain in Georgian (A diacronic analysis)*, Diss. Frankfurt, 2018 (in press).

36. Jakob Kramcritsch, *Geschichte(n) im Netzwerk. Hypertext und dessen Potenziale für die Produktion, Repräsentation und Rezeption der historischen Erzählung*, Münster u.a. 2007.

37. Kalev H. Leetaru, *Culturomics 2.0: Foresasting lange-scale human behavion using global news media tone in time and space*. https://firstmonday.org/article/view/3663/3040

38. Lothar Lemnitzer, Heike Zinsmeister, *Korpuslinguistik. Eine Einführung* (2., durchges. und aktual. Aufl.). Tübingen: Narr., 2010.

39. Wilfried Loth, Wolfgang Wessels, *Auf dem Weg zur Integrationswissenschaft.* In: *Grundlagen für Europa.* Vol. 7, Springer 2001, S. 7–15.

40. Andrea Maihofer, *Inter-, Trans-und Postdisziplinarität. Ein Plädoyer wider die Ernüchterung.* In: Heike Kahlert, Barbara Thiessen, Ines Weller (Hg.): Quer denken – Strukturen verändern. Gender Studies zwischen den Disziplinen. Wiesbaden 2005.

41. Tinatin Margalitadze, *Language@Leeds Working Papers in Linguistics, 1,* University of Leeds. 2020.

42. Tara McPherson. *Introduction: Media Studies and the Digital Humanities* // Cinema Journal. Vol. 48, No. 2, 2009, 119-123. https://www.jstor.org/stable/pdf/20484452.pdf?casa_token=hVBEXMzfuV0AAAAA:UB-DVIx5Jo0NWFl1y2PvLz-WbifpxEQwLZ3EUSc6ZaFDH729UstEUTZ8q5mgim4fdExele7ufkC2TylDwD8YtbM5-cmtVDP_sBqG12yy1cw7WrDmZFil

43. Michael Nentwich, René König, *Cyberscience 2.0. Research in the Age of Digital Social Networks* (= Interaktiva; 11), Frankfurt a.M. 2012.

44. Alexander M. Petersen, Joel Tenenbaum, Shlomo Havlin & H. Eugene Stanley, "*Statistical Laws Governing Fluctuations in Word Use from Word Birth to Word Death*". Scientific Reports. *2:* 313. . (15 March 2012). https://www.nature.com/articles/srep00313

45. Jung-ran Park, *Language-Related Open Archives: Impact on Scholarly Communities and Academic Librarianship*, in.: Electronic Journal of Academic and Special Librarianship. (http://southernlibrarianship.icaap.org/content/v05n02/park_j01.htm)

46. Heinz Pürer, *Publizistik- und Kommunikationswissenschaft. Ein Handbuch.* Konstanz: UVK *2003*, 50.

47. Uwe Quasthoff, Sabine.Fiedler, Erla Hallsteinsdottir (eds.), *Frequency Dictionaty Georgian*, Leipzig, 2018.

48. Carmen Scherer, *Korpuslinguistik.* Heidelberg: Universitätsverlag Winter, 2006.

49. Susan Schreibman, John Unsworth, Ray Siemens, eds., *A New Companion to Digital Humanities.* Oxford: Blackwell, 2016, XVII.

50. Manana Tandaschwili, *Lokation und Deixis im Georgischen - Funktional-semantische Analyse von zed und šig.* In: PRO GEORGIA, Journal of Kartwelological Studies, N. 23, Warschau.

51. Elena Tognini-Bonelli, *Corpus Linguistics at Work*, John Benjamins Publishing, 2001.

52. Arne Ziegler, Karl-Heinz Best, Gabriel Altmann: *A contribution to text spectra*, in: *Glottometrics* 1, 2001, S. 97–108. https://www.ram-verlag.eu/wp-content/uploads/2018/08/g1zeit.pdf#page=100

53. Arne Ziegler, Karl-Heinz Best, Gabriel Altmann: *A contribution to text spectra*, in: *Glottometrics* 1, 2001, S. 98.

ინტერნეტრესურსები:

Civil ენციკლოპედიური ლექსიკონი: http://www.nplg.gov.ge/gwdict/index.php?a=index&d=5

მაიკროსოფტის კორპორაცია: www.microsoft.com

მომხმარებელთა და მოსახლეობის სტატისტიკა: http://www.internetworldstats.com

სოციალური მეცნიერებების ლაიბნიცის ინსტიტუტი: https://www.gesis.org/home

ჰუმანიტარია და სოციალური მეცნიერებები: http://www.h-net.org/

კომუნიკაცია და ინფორმაციები ისტორიულ მეცნიერებაში: http://hsozkult.geschichte.hu-berlin.de/

ფრანკფურტის უნივერსიტეტი, კომპაქტური მეხსიერება: http://www.compactmemory.de

ღია საზოგადოების არქივი: http://www.osaarchivum.org

ვენის უნივერსიტეტის ინფორმატიკის სამსახური: http://pastperfect.univie.ac.at

http://bbf.dipf.de/digitale-bbf/scripta-paedagogica-online

მედიაპორტალი http://www.17juni53.de.

ბერლინის კედლის ქრონიკა: http://www.chronik-der-mauer.de.

11 სექტემბრის დიგიტალური არქივი: http://911digitalarchive.org.

ევროპული ისტორია ონლაინ: http://www.ieg-ego.eu.

ინტერნეტპორტალი Sehepunkte: http://www.sehepunkte.de

ევროპული ისტორიოგრაფიის სარეცენზიო პლატფორმა: http://www.recensio.net

მსოფლიოს ღია ფოტოარქივი: http://www.flickr.com/commons

დოკუპედია: http://docupedia.de

Brandeis-ბიბლიოთეკა: https://blogs.brandeis.edu/library/2012/10/09/whats-digital-humanities-and-how-did-it-get-here/

ტუბინგენის ტექსტის დამუშავების პროგრამების სისტემა: http://www.tustep.uni-tuebingen.de/

https://aa-dh.org/

დიგიტალური ჰუმანიტარიის იაპონიის ასოციაცია: https://www.jadh.org/

დიგიტალური ჰუმანიტარიის ცენტრების მსოფლიო ქსელი: http://dhcenternet.org/

Humanistica: http://www.humanisti.ca/

პერსეუსის დიგიტალური ბიბლიოთეკა: http://www.perseus.tufts.edu/hopper/

გერმანული ლექსიკა 1600 წლის შემდგომ: https://www.dwds.de/

COSMAS II: https://www.ids-mannheim.de/cosmas2/

გეოგრაფიის ინფორმაციული სისტემები: https://www.esri.com/en-us/what-is-gis/history-of-gis

დუბლინის მეტამონაცემების ინიციატივა: Dublin Core Metadata Initiative: http://dublincore.org

ღია არქივების ინიციატივა: Open Archives Initiative: http://www.openarchives.org

ენის ღია რესურსების არქივების გაერთიანება: Open Language Archives Community: http://www.language-archives.org

დიგიტალური ჰუმანიტარიის მანიფესტი: https://www.humanitiesblast.com/manifesto/Manifesto_V2.pdf

მანანა თანდაშვილის საავტორო გვერდი: http://tandaschwili.com/

ინტერნეტჟურნალი „ენა და ლოგიკა": https://geoanbani.com/other/pdfs/Publications/magazine1.pdf

ენის მოდელირების ასოციაცია: http://ena.ge/

უნიკოდის კორპორაცია - ქართული შრიფტების ბლოკები: http://www.unicode.org/charts/PDF/U10A0.pdf
http://www.unicode.org/charts/PDF/U2D00.pdf

სახარების ქართული თარგმანების პარალელური კორპუსი: LOEWE Text collection: NT Novum Testamentum georgice, http://titus.uni-frankfurt.de/texte/etca/cauc/ageo/nt/ntkpl/ntkpl.htm

საქართველოს ეპიგრაფიკული წარწერების კორპუსი: https://iliauni.edu.ge/ge/iliauni/institutebi-451/lingvistur-kvlevata-centri-467/saqartvelos-epigrafikuli-korpusi

Titus-ის პორტალის ქართული რესურსები: http://titus.uni-frankfurt.de/texte/etcg/cauc/ageo/inscr/carcera/carce.htm

ილიას სახელმწიფო უნივერსიტეტის ქართული ენის კორპუსი: http://corpora.iliauni.edu.ge/?q=ka/node/11

ქართული დიალექტების კორპუსი: http://www.corpora.co

ლინგვოკულტუროლოგიური დიგიტალური არქივი: http://digiarchive.bsu.edu.ge

ქართული სამეცნიერო მეტაენის კორპუსი: http://sml.corpora.co/

ქართული პრესის კორპუსი: http://www.opentext.org.ge

საინფორმაციო სააგენტო: http://www.civil.ge

მანჰაიმის ენის კორპუსი: http://www.exmaralda.org

გერმანული ენის საზოგადოება: https://gfds.de/

უნიკოდის ოფიციალური გვერდი: http://www.unicodc.org/versions/Unicode12.0.0/

Codex Sinaiticus: http://www.codex-sinaiticus.net/
Codex Sinaiticus: http://www.codexsinaiticus.org/en/

ლაიფციგის უნივერსიტეტის ტექსტური კორპუსები:
https://corpora.uni-leipzig.de/de?corpusId=kat_mixed_2017&word=#
დიდი ინგლისურ-ქართული ონლაინლექსიკონი www.dict.ge
ინგლისურ-ქართული სამხედრო ონლაინლექსიკონი (http://mil.dict.ge)
ინგლისურ-ქართული ბიოლოგიური ონლაინლექსიკონი" (http://bio.dict.ge)
ინგლისურ-რუსულ-ქართული ტექნიკური ტერმინოლოგიის ონლაინლექსიკონი (https://techdict.ge/)
ინგლისურ-ქართული პარალელური კორპუსი (არიანე ჭანტურიას თარგმანები) http://corp.dict.ge
პაულ მოირერის ქართული ენის კორპუსი: http://clarino.uib.no/korpuskel/corpus
ქართული ლექსიკონი https://www.ganmarteba.ge/
ნოა ბუბენჰოფერის საავტორო გვერდი: http://www.bubenhofer.com/korpuslinguistik/kurs/index.php?id= korpustypen.html
საქართველოს პარლამენტის ბიბლიოთეკის ციფრული ბიბლიოთეკა: https://www.nplg.gov.ge/geo/Digital_Library