

პრეზენტაცია

ირინა ლობჯანიძე

„ვეფხისტყაოსნის“ ონ-ლაინ კორპუსი

შესავალი

„ვეფხისტყაოსანი“ ქართული ლიტერატურის საუკეთესო ნიმუშია. ჩვენამდე მოღწეულია „ვეფხისტყაოსნის“ 162 ხელნაწერი და გამოცემა და 32 ფრაგმენტი. ბეჭდურმა გამოცემამ კი ორმოცდაათს გადააჭარბა. ამჟამად, როცა ვსაუბრობთ „ვეფხისტყაოსანზე“, ვგულისხმობთ არა მხოლოდ მის მხატვრულ ტექსტს, ენობრივ თავისებურებას, არამედ ამ ტექსტის შესახებ შექმნილ სამეცნიერო და პოპულარული ხასიათის მრავალფეროვან ლიტერატურასაც. განსაკუთრებით აღსანიშნავია „ვეფხისტყაოსნის“ უცხოური თარგმანები, მათ შორის ინგლისური, გერმანული, უკრაინული, რუსული და სხვ. ვეფხისტყაოსნის ონ-ლაინ კორპუსი საზოგადოებას საშუალებას აძლევს, ერთმანეთს შეუდაროს აღნიშნული ნაწარმოების არსებული გამოცემები და თარგმანები, შეარჩიოს სასურველი ვერსიები, მიიღოს უფრო სრული წარმოდგენა მათ შესახებ.

ტექსტური მონაცემების დამუშავების ერთ-ერთ შედეგს კორპუსი წარმოადგენს, რომელიც, თავის მხრივ, ემსახურება როგორც ენობრივი მასალის შესწავლას, ასევე მის გამოყენებას სხვადასხვა მიზნისათვის, როგორცაა მაგალითად ლექსიკოგრაფია და გამოყენებითი ლინგვისტიკა. ბუნებრივია, ისმის რამდენიმე კითხვა: 1. როგორი უნდა იყოს მსგავსი კორპუსის ტექსტონომია; 2. რა სხვაობაა ნაბეჭდ გამოცემასა და ხელნაწერს შორის მეტა-ანოტირების მხრივ; და, საბოლოოდ, 3. რას უნდა ემყარებოდეს ბალანსირების პრინციპი.

საკვლევი საკითხი

კორპუსის ლინგვისტიკაში მიღებული ზოგადი დეფინიციების შესაბამისად, კორპუსი წარმოადგენს ტექსტთა ერთობლიობას, რომელიც ერთიანდება განსაზღვრული პარამეტრებისა და პრინციპების გათვალისწინებით. ფართო გაგებით, კორპუსის განმარტება წარმოგვიდგება შემდეგი ამონარიდების სახით:

1. კორპუსი წარმოადგენს ბუნებრივ ენობრივ ტექსტთა კრებულს, რომელიც შერჩეულია ენობრივი მდგომარეობისა ან მრავალფეროვნების ასახვის მიზნით (Sinclair 1991, 171);

2. კორპუსი გაიგივებულია ჩვეულებრივი ტექსტური ამონარიდების კოლექციასთან. ტექსტები შეიძლება იყოს ზეპირი, წერილობითი ან შერეული სახის, ხოლო მაგალითები – ნებისმიერი სიგრძის (Aarts 1991, 45).

ზემოხსენებული განმარტებების გარდა, კორპუსის ლინგვისტიკა მიმართავს სხვა საშუალებებსაც აღნიშნული ტერმინების განსაზღვრის მიზნით, მაგრამ დღემდე არ არსებობს საბოლოოდ შეჯერებული და ყველასათვის მისაღები დეფინიცია, რომელსაც ავტორთა უმრავლესობა იზიარებდეს. უფრო მეტიც, სხვადასხვა ენის კორპუსებში წარმოდგენილი პარამეტრები განსხვავდება როგორც ლინგვისტური, ასევე ტექნიკური თვალსაზრისითაც. შედარების მიზნით, შეგვიძლია, გადავხედოთ ინგლისურ, ახალბერძნულ, რუსულ ან ნებისმიერ სხვა კორპუსებს¹ და, ვნახავთ, რომ საერთო თვისებების მიუხედავად, თითოეულ კორპუსს განსხვავებული პარამეტრებიც ახასიათებს. და, შესაბამისად, კორპუსთა უმრავლესობა ემსახურება არა რომელიმე ერთი კონკრეტული ამოცანის გადაწყვეტას, არამედ ამოცანათა კომპლექსს. ამის მიხედვით, კორპუსის შექმნის წინაპირობად მიჩნეულია შემდეგი დებულებები:

1. უნდა არსებობდეს წინასწარ განსაზღვრული ამოცანა, რომლის გადაწყვეტას ემსახურება ამა თუ იმ კორპუსის შექმნა – ჩვენ შემთხვევაში, კორპუსი ჩაფიქრებული იყო, როგორც ინტერდისციპლინარული, მულტილინგვური პროექტი, რომლის მიზანია „ვეფხისტყაოსნის“, როგორც მსოფლიო კულტურული მემკვიდრეობის მნიშვნელოვანი ძეგლის მრავალმხრივი გააზრება და ამ პოემასთან დაკავშირებული მონაცემების შეგროვება, სისტემატიზაცია, გამოქვეყნება და კვლევა;

1 ინგლისური ენის ეროვნული კორპუსი – <http://www.natcorp.ox.ac.uk/>, ახალი ბერძნული ენის ეროვნული კორპუსი – <http://hnc.ilsp.gr/en/>, რუსული ენის ეროვნული კორპუსი – <http://www.ruscorpora.ru/en/index.html> და სხვ.

2. უნდა არსებობდეს კორპუსის წინასწარ განსაზღვრული დიზაინი – ჩვენ შემთხვევაში, პორტალზე იგეგმებოდა რამდენიმე ძირითადი ბლოკის განთავსება: ა) პარალელური კორპუსის ბლოკი (რომელიც გულისხმობდა ყველა გამოცემის, ხელნაწერის და არსებული უცხოენოვანი თარგმანის კვლევისა და შედარების საშუალებას); ბ) ვიზუალური მემკვიდრეობის ბლოკი (რომელიც შეიცავს თითოეული ხელნაწერიდან, როგორც მინიმუმ, ერთ გვერდს და, ბუნებრივია, აძლევს ჩვეულებრივ მომხმარებელს მინიატურების, ყდისა და კალიგრაფიის შედარების შესაძლებლობას); გ) სამეცნიერო ლიტერატურის ბლოკი (რომელიც გულისხმობს პოემის ირგვლივ არსებული სამეცნიერო ლიტერატურის მოძებნის საშუალებას);

3. უნდა არსებობდეს ტექსტების კრებული – ჩვენ შემთხვევაში, იგულისხმებოდა ქართულენოვანი და ინგლისურენოვანი გამოცემების ელექტრონული ვერსიების შექმნა ან შეგროვება და, რასაკვირველია, ხელნაწერთა ეროვნულ ცენტრში დაცული ხელნაწერების დიגיტიზაცია განსაზღვრული პარამეტრების მიხედვით;

4. უნდა არსებობდეს პრინციპი და კრიტერიუმი ამ ტექსტების შერჩევისა;

5. უნდა არსებობდეს ტექსტების ელექტრონული ფორმა (machine-readable standard).

ძირითადი მიზანი, რადგანაც უკავშირდებოდა ონ-ლაინ კორპუსის შექმნას, გამოსაყენებელი და გასათვალისწინებელი იყო ზოგიერთი სტანდარტი, კერძოდ, 1. ბუნებრივი ენის დამაშუვების ISO-ს ძირითადი სტანდარტები: სეგმენტაცია (ISO 24614), ანოტაცია (ISO 24611, 246121 და 24615), მახასიათებლები (ISO 24610), ლექსიკონები (ISO 24613); 2. დამატებითი სტანდარტები: მონაცემთა კატეგორიები (ISO 12620), ენობრივი კოდები (ISO 639 ან IETF BCP-47), სკრიპტ-კოდები (ISO 15924), ქვეყნის კოდები (ISO 3166), თარიღი (ISO 8601) და უნიკოდი (ISO 10646) [ISO-TC37] და TEI-ს რეკომენდაციები.

შესაბამისად, კვლევის ძირითად ამოცანად გვევლინებოდა შემდეგი საკითხები: ა) ტექსტების დამუშავება და მეტა-ანოტირების პრინციპი; ბ) ბალანსირების პრინციპი; გ) კორპუსის დიზაინი და ტაქსონომია.

მეტა-ანოტირების პრინციპი

მეტა-ანოტირება გვაძლევს დამატებით მონაცემებს კორპუსში განთავსებული ტექსტების შესახებ, სხვაგვარად, რომ ვთქვათ, მეტა მონაცემები განისაზღვრება, როგორც „მონაცემები მონაცემთა შესახებ“ (Bournaud 2005, 30). ენობრივ კორპუსთა უმრავლესობაში, რო-

გორც წესი, გვხვდება სხვადასხვა ტიპის მეტა-ანოტირება როგორც საკმაოდ მარტივი, ასევე საკმაოდ რთული სქემებით. უზოგადესი სახით, შეიძლება შეგვხვდეს შემდეგი: ა) ბიბლიოგრაფიული მონაცემები კორპუსში ჩადებული ნაბეჭდი ტექსტის შესახებ; ბ) აღწერითი ხასიათის მონაცემები კორპუსის კომპონენტების შესახებ; გ) დოკუმენტური ხასიათის მონაცემები თვითონ კორპუსის შესახებ. აღნიშნული ტიპის მონაცემების განსაზღვრაზე ბოლო წლების განმავლობაში დიდ ზეგავლენას ახდენს TEI-ს სტანდარტი.

ვეფხისტყაოსნის კორპუსში მოსათავსებელი იყო პოემის როგორც ნაბეჭდი ტექსტი, ასევე ხელნაწერი და, ბუნებრივია, რომ მათი მეტა-ანოტირების სქემაც განსხვავებულია, რადგანაც გვხვდებოდა:

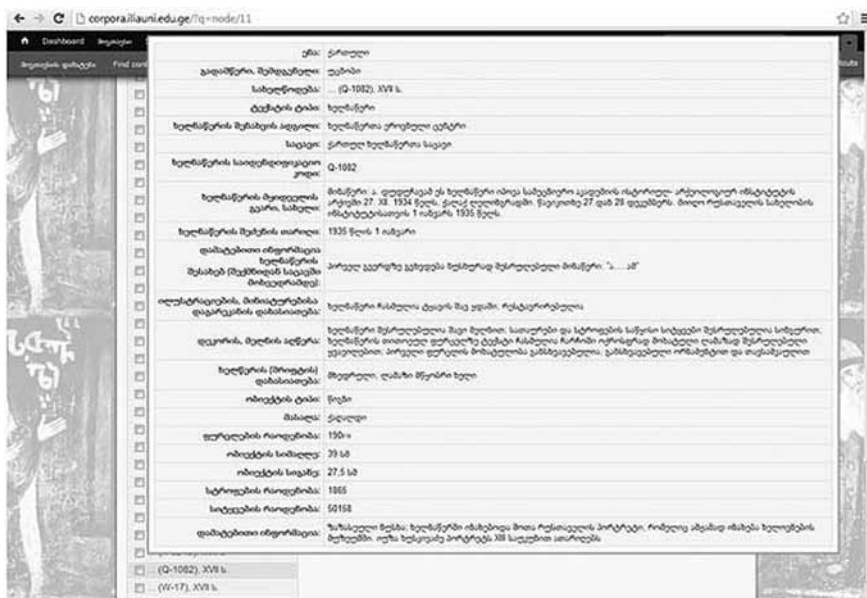
ნაბეჭდი ტექსტი, რომლის საფუძველს წარმოადგენდა კონკრეტული ხელნაწერი ან ხელნაწერები;

ნაბეჭდი ტექსტი, რომლის საფუძველს წარმოადგენდა უკვე გამოცემული ტექსტი;

ხელნაწერი, რომელიც არსად გამოცემულა და, საბოლოოდ;

ხელნაწერი, რომელიც უკვე იყო გამოცემული.

მსგავს შემთხვევებში მეტა-ანოტირების სქემები ერთმანეთისგან განსხვავდებოდა. სხვაობა კი მათ შორის უკავშირდებოდა, ისეთ დამატებით ინფორმაციას ხელნაწერების შესახებ, როგორიცაა ა) გადამწერისა ან შემდგენელის ვინაობა; ბ) შენახვის ადგილი; გ) საცავის ნომერი; დ) ობიექტის ზომა და ტიპი; ე) მასალა; ვ) დამატებით ინფორმაციას ხელნაწერის შესახებ მისი შექმნიდან საცავში მოხვედრამდე და ა.შ. აღნიშნული მეტა-მონაცემების ვიზუალური ასახვა შემდეგნაირია:



აღწერილობიდან ვხვდებით, რომ ვეფხისტყაოსნის ხელნაწერის Q-1082 გადამწერი უცნობია, ხოლო თვით ხელნაწერი შეიქმნა XVII საუკუნეში, ხოლო ნაპოვნი იყო სამეცნიერო აკადემიის ისტორიულ-არქეოლოგიური ინსტიტუტის არქივში 1935 წელს. მინაწერის ფორმატიც დაცულია ზუსტად ისე, როგორც გვხვდება ხელნაწერში.

ძირითადი ინფორმაცია, რომელიც მოსდევს კორპუსში ჩადებულ ყველა ფაილს შემდეგია:

თვისება (Feature)	(Value)	ტაგები XML ფაილში
პროექტის დახასიათების ბლოკი		
დამფინანსებელი ორგანიზაცია	ტექსტი	<funder>
წამყვანი ორგანიზაცია	ტექსტი	<principal>
პასუხისმგებელი პირის სახელი, გვარი	ტექსტი	<name>
პასუხისმგებელი პირის ფუნქცია	ტექსტი	<resp>
პასუხისმგებელი პირი, ორგანიზაცია	ტექსტი	<respStmt>
პროექტის სახელწოდება	ტექსტი	<projectDesc>
ფაილის დახასიათების ბლოკი		
ფაილის ავტორის გვარი, სახელი	ტექსტი	<name>
ფაილის წყარო	ტექსტი	<sourceDesc>
ფაილის ენა	kat, en, fro, rus, tur	<langUsage>

თვისება (Feature)	(Value)	ტაგები XML ფაილში
ფაილის ზომა, კბ	[ავტომატური დათვლა]	<measure quantity=„ unit=„kB„>
ფაილის შექმნის თარიღი	ციფრი	<date value=„2010-2020„>
ფაილის შექმნის ადგილი	ტექსტი	<rs type=„city„>
ინფორმაცია ფაილში შეტანილი ცვლილებების შესახებ	ტექსტი	<revisionDesc>
1 ცვლილების თარიღი	ციფრი	<date>
1 ცვლილების ტიპი	ტექსტი	<change>
1 კორექტორის გვარი, სახელი	ტექსტი	<name>
2 ცვლილების თარიღი	ციფრი	<date>
2 ცვლილების ტიპი	ტექსტი	<change>
2 კორექტორის გვარი, სახელი	ტექსტი	<name>
სტროფების რაოდენობა	[ავტომატური დათვლა]	<measure type=„stanzas“ quantity=„„>
სიტყვების რაოდენობა	[ავტომატური დათვლა]	<measure type=„words“ quantity=„178869„>
ნაბეჭდი ტექსტის დახასიათების ბლოკი		
ტექსტის სახელწოდება	ტექსტი	<title>
ავტორის სახელი, გვარი	ტექსტი	<author>
წყარო ენა	kat, en, fro, rus, tur	<textLang>
ტექსტის შექმნის თარიღი	ციფრი: წელი, სა- უკუნე	<date when=“1937“></ date>
ტექსტის შექმნის ადგილი	ტექსტი	<origPlace>
გამომცემელი	ტექსტი	<publisher>
გამოცემის ადგილი	ტექსტი	<pubPlace>
გამოცემის თარიღი	ციფრი	<date when=„1937„></ date>
რედაქტორი	ტექსტი	<editor>

თვისება (Feature)	(Value)	ტაგები XML ფაილში
მთარგმნელი	ტექსტი	<editor role=„translator„>
ილუსტრატორი	ტექსტი	<editor role=„illustrator„>
ტომების/ნომრების რაოდენობა	ციფრი	<idno type=„vol“ n=„1.2“ />
გვერდების რაოდენობა	ციფრი	<extent>
ტექსტის გვერდები – დან – მდე	ციფრი	<extent from=„1“ to=„2“ />
ISBN/ISSN	ციფრი	<idno type=„ISBN“ />
წვდომა	თავისუფალი, შეზღუდული, სხვა	<availability>
გამავრცელებელი	ტექსტი	<distributor>
უფლებამოსილი ორგანიზაცია	ტექსტი	<authority>
შენიშვნები	ტექსტი	<note>
ხელნაწერის დახასიათების ბლოკი		
შენახვის ადგილი	ტექსტი	<settlement>
საცავის სახელწოდება	ტექსტი	<repository>
საცავის ნომერი	ტექსტი	<idno>
კოლექციის სახელწოდება	ტექსტი	<collection>
დამატებითი საიდ. კოდი, კატალოგის ნომერი	ტექსტი	<altIdentifier>
ხელნაწერის ავტორის, გადამწერის ან შემდგენელის გვარი, სახელი	ტექსტი	<name>
ხელნაწერის ავტორის ფუნქცია	შემდგენელი, გადამწერი, მთარგმნელი, სხვა	<editor role=„compiler„>
ხელნაწერის სახელწოდება	ტექსტი	<title>
ხელნაწერის ენა და საწერი სისტემა		

თვისება (Feature)	(Value)	ტაგები XML ფაილში
ხელნაწერის ენა და საწერი სისტემა (ასომთავრული, ნუსხური, მხედრული)	ტექსტი	<textLang mainLang=„ka“ otherLangs=„HEL„>ენა: ძველი ქართული; შრიფტი: მხედრული</textLang>
ხელნაწერის ფიზიკური მდგომარეობის აღწერა		
ობიექტის ფორმა	კოდექსი, გრაგნილი, რვეული, ფრაგმენტი, ფურცელი, სხვა	<objectDesc form=„codex„>
მასალა, ქალაღი	ჭილი, ეტრადი, ქალაღი	<supportDesc material=„paper„>
ფურცლების რაოდენობა	ტექსტი	<extent>
ფურცლის ზომის ტიპი	ტიპი: ფურცელი; ერთეული: სმ, მმ	<dimensions scope=„all“ type=„leaf“ unit=„inch„>
სიმაღლე	ციფრი	<height>
სიგანე	ციფრი	<width>
ხელნაწერის მდგომარეობა (ცვლილებების, დაზიანებების აღწერა)	ტექსტი	<condition>
ხელნაწერის ფურცლების დანომრვის აღწერა (მაგ., რექტო, ვერსო და ა.შ.)	ტექსტი	<foliation>
ფურცლების თანმიმდევრობის დახასიათება (მაგ. არეული თანმიმდევრობა და ა.შ.)	ტექსტი	<collation>
ხელნაწერის ფორმის აღწერა		
ხელის აღწერა	ტექსტი	<handDesc>

თვისება (Feature)	(Value)	ტაგები XML ფაილში
სკრიპტის აღწერა	ტექსტი	<scriptDesc>
მინიატურებისა და დეკორაციების აღწერა	ტექსტი	<decoDesc>
მინაწერები	ტექსტი	<additions>
ისტორია		
ხელნაწერის შექმნის ადგილი	ტექსტი	<origin>
ხელნაწერის შექმნის თარიღი (date) – დან	ციფრი	<origDate notAfter=„1600“ notBefore=„1700„>
ხელნაწერის შექმნის თარიღი (date) – მდე	ციფრი	<origDate notAfter=„1600“ notBefore=„1700„>
ინფორმაცია ხელნაწერის შესახებ საცავში მოხვედრამდე (არსებობის შემთხვევაში)	ტექსტი	<provenance>
ინფორმაცია ხელნაწერის შეძენისა ან ჩუქების შესახებ	ტექსტი	<acquisition>

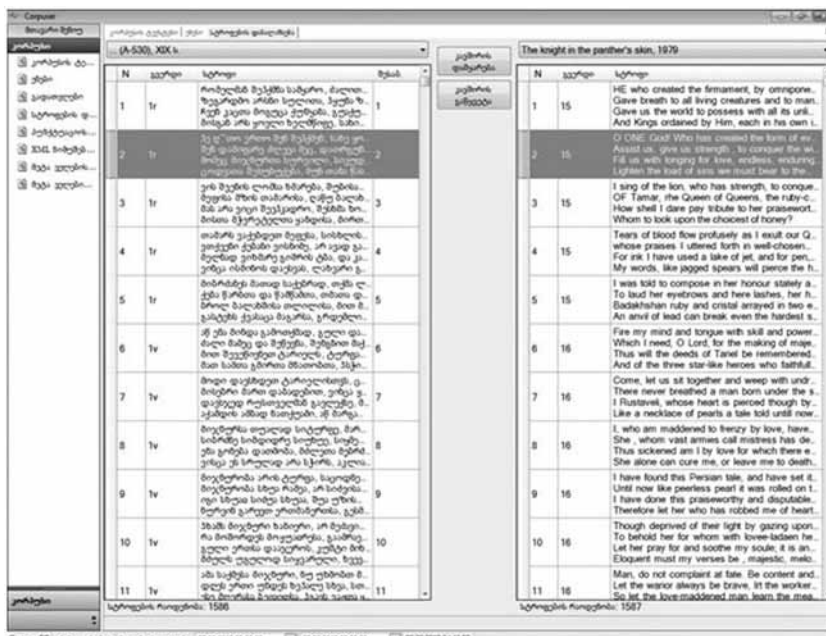
ბუნებრივია, რომ აღნიშნული ინფორმაციის მოგროვება და მისი დამუშავება საკმაოდ ხანგრძლივი და შრომატევადი პროცესია. განსაკუთრებით მაშინ, როცა საკითხი უკავშირდება გამოუცემელ ხელნაწერებს. ამიტომაც, ხელნაწერების ტექსტების დიგიტალიზაციაზე მუშაობდნენ ცენტრის თანამშრომლები ც. ხახვიაშვილი, ნ. ბილანიშვილი, გ. შუბითიძე და ლ. კუპრეიშვილი. მათ უწევდათ როგორც ხელნაწერების, ასევე მინაწერების ამოკითხვა და საჭირო ფორმატამდე დაყვანა.

ბალანსირების პრინციპი

კორპუსის შექმნის პროცესში, საჭირო გახდა არანაკლებ მნიშვნელოვანი საკითხის გადაწყვეტა, როგორიცაა ტექსტების დაბალანსება. პარალელური კორპუსის მთავარი პრინციპი ერთსა და იმავე ტექსტის წარმოდგენა როგორც ორიგინალის ენაზე, ასევე უცხო ენაზეც. ბუნებრივია, არ არის საკმარისი მხოლოდ პარალელური ტექსტების არსებობა, პარალელური კორპუსი უნდა იყოს განსაზღვრულ დონეზე გასწორებული, მაგალითად, წინადადების ან სიტყვის დონეზე (McEnery 2011, 70). გასწორება, რომელიც, ძირითადად, ავტომატურად სრულ-

დება, შეიძლება შეგხვდეს როგორც წინადადების, ასევე სტროფის დონეზეც, როცა საკითხი ეხება პოემას.

ხელნაწერებისა და ნაბეჭდი ტექსტების ერთმანეთთან დაბალანსების უზრუნველყოფისათვის არსებობდა საეტალონო ტექსტის განსაზღვრის საშუალება, რომელიც გულისხმობდა პოემის ყველაზე სავესტ ტექსტის განსაზღვრას და მისი სტროფების ამოსავალ ერთეულებად გამოყენებას ყველა დანარჩენი ტექსტისთვის. საეტალონო ტექსტის განსაზღვრა შეუძლებელი აღმოჩნდა, რადგანაც გამოცემები და ხელნაწერები ერთმანეთისაგან განსხვავდებოდა როგორც სტროფების რაოდენობის მხრივ, ასევე შინაარსობრივადაც. ამიტომაც პროექტის მსვლელობისას შეიქმნა კომპიუტერული უზრუნველყოფა Corpuser-ი (პროგრამისტი: მერაბ ზაკალაშვილი).



აღნიშნული პროგრამა ნახევრადავტომატურ რეჟიმში გვაძლევდა გარე-ანოტირებისა და სტროფების დაბალანსების შესაძლებლობას საეტალონო ტექსტის გარეშე. პროგრამაში ერთმანეთის პარალელურად შესაძლებელი იყო ქართული და ინგლისური ტექსტის ნახვა და კავშირების დამატებითი ჩასწორება. თუმცა, ვებ-გვერდებზე პროგრამის შედეგების გამოტანის პრინციპი განსხვავებულია.

კორპუსის დიზაინი და ტექსონომია

ნებისმიერი კორპუსის მთავარ დანიშნულებას ის ინფორმაცია წარმოადგენს, რომლის ნახვა შეუძლია მის ძირითად მომხმარებელს. ამ მხრივ, ძალზე მნიშვნელოვანი იყო თავიდანვე იმისი განსაზღვრა თუ როგორი ტიპის ინფორმაციის ნახვა იქნებოდა შესაძლებელი ონ-ლაინ სივრცეში. თუმცა პროექტის მსვლელობისას აღნიშნული ინფორმაცია ექვემდებარებოდა რიგ ცვლილებას. ამ ეტაპზე, ვეფხისტყაოსნის პორტალი მოიცავს შემდეგი სახის ძირითად ინფორმაციას:

1. ვეფხისტყაოსნის პარალელური კორპუსის ბლოკს, რომელიც ფაქტიურად ორ ამოცანას ემსახურება: 1. გვაძლევს პოემის ყველა ქართულენოვანი ბეჭდური გამოცემისა და საქართველოში დაცული ხელნაწერების ელექტრონული ვერსიების ნახვისა და წაკითხვის შესაძლებლობას; 2. გვაძლევს პოემის ყველა ქართულენოვანი გამოცემის, ხელნაწერებისა და ინგლისურენოვანი თარგმანების დაბალანსებული და სტროფის დონეზე გასწორებული ტექსტების კვლევის შესაძლებლობას. ამ უკანასკნელში შეტანილია პოემის ინგლისურენოვანი თარგმანები – სულ ხუთი გამოცემა, მათ შორის მარჯორი უორდროპის და სტივენსონის. ბლოკში ძიება შესაძლებელია როგორც ცალკეული გამოცემების, ხელნაწერებისა და ინგლისური თარგმანების, ასევე ყველა არსებული ტექსტური რესურსის მიხედვით.

2. სამეცნიერო ლიტერატურის ბლოკში წარმოდგენილია პოემის ირგვლივ არსებული ძირითადი სამეცნიერო ლიტერატურა. აღნიშნული ბლოკი დღემდე ივსება სამეცნიერო ხასიათის ლიტერატურით.

3. ვიზუალური მემკვიდრეობის ბლოკი მომზადდა ხელნაწერთა ეროვნულ ცენტრთან მჭიდრო თანამშრომლობით. მასში განთავსდა თითო გვერდი ილუსტრირებული ხელნაწერებიდან, ასევე კალიგრაფიისა და ყდის ნიმუშები.

4. ხელნაწერთა მინაწერების ბლოკში წარმოდგენილია სხვადასხვა ეპოქის ხელნაწერებზე დართული მეტატექსტები. მათი უმეტესობა პირველად ქვეყნდება.

დამოწმებანი

- Aarts, J.**, 1991. "Intuition-based and observation-based grammars" English Corpus Linguistics. Studies in Honour of Jan Svartvik, 44-62
- Bournaud, Lou**, 2005. "Metadata for Corpus Work" B Developing Linguistic Corpora: A Guide to Good Practice, M. Wynne, Oxford: Oxford Books, 30-46
- McEnery, T., Wilson, A.**, 2011. Corpus Linguistics. Edinburgh: Edinburgh University Press
- Meurer, P.**, 2011. "Constructing an annotated corpus for Georgian – Tools and resources" Kutaisi Symposium on Language, Logic and Computation, Batumi
- Sinclair, J.**, 1991. Corpus, concordance, collocation: Describing English Language. Oxford: Oxford University Press
- Tognini-Bonelli, E.** 2001., Corpus linguistics at work. Amsterdam: Benjamins.

PRESENTATION

Irina Lobzhanidze

The Online Corpus of *The Knight in the Panther's Skin*

The Knight in the Panther's Skin can be considered as the most brilliant Georgian poem. There are about 162 manuscripts and printed editions as well as 32 fragments of the poem. When we talk about *The Knight in the Panther's Skin*, we mean not only a literary text and its linguistic features, but also rich material for cultural studies. The poem has been translated into many languages including English, German, Ukrainian and Russian.

The online corpus of *The Knight in the Panther's Skin* enables the public to compare the existing editions and translations of the poem, to select the desired versions and to receive complete information about them.