

საქართველოს პარლამენტი

„ქართული ენა - 21-ე საუკუნის
გამოწვევები“
კონფერენციის მასალები

თბილისი
2013

წინამდებარე კრებულში წარმოდგენილია 2013 წლის 8 ივლისს, ქ. ქუთაისში, საქართველოს პარლამენტის განათლების, მეცნიერებისა და კულტურის კომიტეტის ინიციატივით ჩატარებული კონფერენციის - „ქართული ენა - 21-ე საუკუნის გამოწვევები“ - მასალები.

კონფერენციის სესიები მიემდგვნა ქართული ენის ინსტიტუციონალური მხარდაჭერისა და განვითარების პერსპექტივებს, ქართული ენის ეროვნული კორპუსის შექმნის მნიშვნელობას, ქართული ენის ჩართულობას თანამედროვე ტექნოლოგიებში.

კონფერენციის მუშაობაში მონაწილეობა მიიღო 50-მდე მეცნიერმა, აგრეთვე საქართველოს პარლამენტის წევრებმა, აღმასრულებელი ხელისუფლების წარმომადგენლებმა, ექსპერტებმა.

გამოცემის რედაქტორი - ფრიდონ საყვარელიძე

ქართული ენის მანქანური თარგმანის პროგრამა - პრობლემები და პერსპექტივები

რატი სხირტლაძე

იუნესკოს შეფასებით, 2100 წლისთვის, დღეს არსებული ენების 90% შესაძლოა ჩანაცვლდეს დომინანტური ენებით. აღსანიშნავია, რომ ენების 96%-ზე მსოფლიოს მოსახლეობის მხოლოდ 3% საუბრობს¹. გლობალიზაციის არსებული ტემპის გათვალისწინებით, სიტუაცია მართლაც საგანგაშოა.

დღეისათვის ინტერნეტში არსებული ინფორმაციის უდიდესი ნაწილი ინგლისურენოვანია და არის მოსაზრება, რომ ინგლისური გახდება მომავლის ენა (საერთაშორისო დამხმარე ენა, ე.წ. *ინტერლინგვა*). ინტერნეტში ქართულენოვანი ინფორმაციის წილი დღეს მხოლოდ 0.006%-ია² (შედარებისთვის - რუსული - 6.4%, თურქული - 1.3%, ლიტვური - 0.1%. აღსანიშნავია, რომ ეს უკანასკნელი ქართულენოვან მაჩვენებელს 17-ჯერ აღემატება).

რა პერსპექტივა აქვს ქართულ ენას გლობალიზაციის პირობებში?

თავისთავად მნიშვნელოვანია, რომ ქართულ ენას სახელმწიფო სტატუსი აქვს, მაგრამ ვფიქრობთ, რომ ესეც არ არის სრული გარანტია იმისა, რომ 21-ე საუკუნის გამოწვევების გათვალისწინებით, როგორც სრულფასოვან ენას, მას საფრთხე არ ემუქრება. სავსებით რეალურია, რომ ციფრულ ერაში იგი დარჩეს მხოლოდ ოფიციალური, ოჯახურ და ნაცნობ-მეგობრებთან კომუნიკაციის ენად, მაგრამ დაკარგოს სამუშაო, საგანმანათლებლო და ინფორმაციის მისაღებად გამოყენებული ენის ფუნქცია.

რა არის გამოსავალი?

ქართულად მოსაუბრე ადამიანების სიმცირის გამო იმედი არ უნდა ვიქონიოთ, რომ მათ მიერ შექმნილი ქართულენოვანი ინფორმაცია (კონტენტი) დააკმაყოფილებს ინფორმაციაზე მზარდ მოთხოვნილებას.

მთარგმნელების დახმარებით თუნდაც მხოლოდ კრიტიკული მნიშვნელობის უცხოენოვანი ინფორმაციის ქართულ ენაზე და ქართულიდან უცხო ენებზე თარგმნა ძალზედ ძვირი სიამოვნებაა (მაგალითისთვის, ვეროკავშირის ინსტიტუციები

1 UNESCO Ad Hoc Expert Group on Endangered Languages (2003). "Language Vitality and Endangerment" (pdf).

2 <http://w3techs.com/technologies/details/cl-ka/all/all>

თარგმანზე წელიწადში დაახლოებით 1 მლრდ ევროს ხარჯავენ¹⁾ და შესასრულებელი სამუშაოს მოცულობიდან გამომდინარე, უტოპიურია.

ამ სიტუაციაში პრაქტიკულად ერთადერთი გამოსავალი მანქანური თარგმანის პროგრამის შექმნაა, რომელიც ამ პრობლემებს მოაგვარებს.

რა მდგომარეობა მსოფლიოში მანქანური თარგმანის მიმართულებით?

აღსანიშნავია, რომ ინტერნეტის მომხმარებელთა 70% ინგლისურს არ ფლობს²⁾. გლობალიზაციამ მკვეთრად გაზარდა ინგლისურ ენაზე მოთხოვნა, მაგრამ მედალს მეორე მხარეც აქვს - გლობალურმა ბიზნეს კომპანიებმა დაინახეს, რომ ერთენოვანი მარკეტინგი დღევანდელ მოთხოვნებს ვეღარ უმკლავდება³⁾ და ბიზნესის განვითარებისთვის მრავალენოვანი მარკეტინგია აუცილებელი. გლობალური ბიზნესი ვეღარ ელოდება იმ დროს, როცა მსოფლიო ერთ ინტერლინგვაზე აღაპარაკდება. შესაბამისად, დღეს, როგორც არასდროს, განსაკუთრებით მძაფრად დაისვა მანქანური თარგმანის პროგრამების შექმნის საკითხი და ისეთი მასშტაბის კომერციული კომპანიები, როგორცაა გუგლი და მაიკროსოფტი, რესურსებს არ იშურებენ ამ პრობლემის გადასაწყვეტად.

მანქანური თარგმანის პროგრამის შექმნაზე მუშაობა ჯერ კიდევ გასული საუკუნის 50-იანი წლებიდან დაიწყო. ამოცანა ძალზედ კომპლექსურია და დღემდე ღია პრობლემად რჩება, თუმცა, ბოლო წლებში სახეზეა არსებითი წინსვლა. მანქანური თარგმანის პრობლემის გადასაწყვეტად არსებობს ორი ძირითადი მიდგომა: წესებზე დაფუძნებული მანქანური თარგმანი და სტატისტიკაზე (ე.წ. „დაგროვების პრინციპით“) დაფუძნებული მანქანური თარგმანი.

სტატისტიკაზე დაფუძნებული მანქანური თარგმანი დღეისათვის წესებზე დაფუძნებულ მანქანურ თარგმანზე მეტად ეფექტურად მიიჩნევა. ეს გზა აირჩია Google-მა, თუმცა მან 2010 წელს მეტად მნიშვნელოვანი განცხადება გააკეთა⁴⁾: რამდენიმე ენისათვის მან

1 <http://info.moravia.com/blog/bid/96247/Machine-Translation-in-the-European-Union>

2 <http://www.milengo.com/blog/marketing-localization-speak-the-language-get-the-customer/>

3 <http://www.milengo.com/blog/preserving-languages-with-machine-translation/>

4 <http://www.andovar.com/blog/machine-translation/>

დაამუშავა არსებული ბილინგვური ტექსტების 10% და შექმნა სტატისტიკაზე დაფუძნებული მანქანური თარგმანის შესაბამისი პროგრამები, მაგრამ დამუშავებული ტექსტების გაორმაგების შემდეგ, თარგმანის ხარისხი მხოლოდ 0.5%-ით გაიზარდა, რაც იმაზე მიუთითებს, რომ თარგმანის ხარისხის გასაუმჯობესებლად საკმარისი არაა სტატისტიკური მეთოდების გამოყენება. გარდაუვალია წესებზე დაფუძნებული სათარგმნი მოდულების შექმნა.

საქართველოში მანქანური თარგმანის პირველი ჯგუფი ჩამოყალიბდა 1957 წელს საქართველოს მეცნიერებათა აკადემიის ელექტრონიკის, ავტომატიკისა და ტელემექანიკის ინსტიტუტში არჩილ ელიაშვილის ხელმძღვანელობით. აღნიშნულ საკითხზე მომდევნო პერიოდში მუშაობდნენ გიორგი ჩიკოიძე, ბესარიონ გუგუშვილი, ლიანა ლორთქიფანიძე და სხვანი.

ვხედავდით რა ქართული ენის მანქანური თარგმანის პროგრამის შექმნის გარდაუვალ აუცილებლობას, ბატონი თედო უთურგაიძის ხელმძღვანელობით, ენათმეცნიერებმა, მათემატიკოსებმა და პროგრამისტებმა ჩამოვყალიბეთ ინტერდისციპლინარული ჯგუფი, რომელმაც, თანამედროვე ტექნოლოგიების გამოყენებით, ქართული ენის მოდელირება დაისახა მიზნად. 2006 წელს, ამ ჯგუფის საფუძველზე, დაფუძნდა „ენის მოდელირების ასოციაცია“.

სამუშაოების პირველ ეტაპზე მიზნად დავისახეთ თედო უთურგაიძის ენობრივი მოდულების შესაბამისად მოგვეზინა ქართული ენის მორფოლოგიური და ნაწილობრივ სინტაქსური მოდელის აგება. შევქმენით ქართული სალიტერატურო ენის პროცესორი. პროცესორის საშუალებით შესაძლებელია მორფემებად დაიშალოს სიტყვაფორმა, დადგინდეს მისი გრამატიკული კატეგორია (რომელი მეტყველების ნაწილია, სახელის შემთხვევაში - რომელ ბრუნვაშია, ზმნის შემთხვევაში - უღლების რომელი კატეგორიაა და ა.შ). აგებულია 592 ზმნური და 72 სახელური მოდელი, რომლებიც, შესაბამისად, 29 მლნ და 109 მლნ, სულ 138 მლნ განსხვავებულ სიტყვაფორმას აწარმოებენ.

მართალია, ქართული სალიტერატურო ენის პროცესორის შექმნის უმთავრესი მიზანი სათარგმნი პროგრამისთვის ბაზისის მომზადება იყო, მაგრამ იგი საშუალებას გვაძლევს, შევქმნათ ძალიან ეფექტური მართლწერის შემმოწმებელი პროგრამაც. ფონდი „ღია საზოგადოება - საქართველოს“ მხარდაჭერით შეიქმნა „მაიკროსოფტ ვორდთან“ ინტეგრირებული ქართული ენის მართლწერის

შემმოწმებელი პროგრამა. პროგრამის სამომხმარებლო ხასიათმა და ქართული სიტყვაფორმების უპრეცედენტო სიმრავლემ ახალი გამოწვევების წინაშე დაგვყენა - აუცილებელი გახდა პროცესორის სისწრაფის გაზრდა, რამაც, თავის მხრივ, ქართული ენის არსებული მოდელის მნიშვნელოვანი ცვლილებები მოითხოვა. შესაბამისი ცვლილებების შემდეგ, ახალი საპილოტე პროგრამის სისწრაფე წინამორბედზე დაახლოებით 80-ჯერ გაიზარდა.

რუსთაველის ფონდის დახმარებით, ენის მოდელირების ასოციაციამ შექმნა ქართული ენის ორთოგრაფიული ლექსიკონი, რომელიც 170,000-ზე მეტ ერთეულს შეიცავს და რომლის ნახვაც ასოციაციის ვებგვერდზეა შესაძლებელი (www.ena.ge). მან მოხმარებლებში ფართო პოპულარობა მოიპოვა, რასაც მოწმობს ჩვენს ფორუმზე განთავსებული წერილები. რუსთაველის ფონდის დახმარებით, ენის მოდელირების ასოციაციის და ენათმეცნიერების ინსტიტუტის ერთობლივი ძალისხმევით, ასევე შეიქმნა ქართული ენის განმარტებითი ლექსიკონის ელექტრონული ვერსიაც (www.ena.ge). იგი კომპლექსური საძიებო სისტემითაა აღჭურვილი, რაც არამარტო ფართო საზოგადოებაზე, არამედ ლექსიკოგრაფიული კვლევების ჩატარებაზეა ორიენტირებული. იგი მკვლევარს საშუალებას აძლევს კონკრეტული დიალექტის, დარგის, მეტყველების ნაწილის თუ სხვა მრავალი პარამეტრის შესაბამისად მოიძიოს და დაამუშაოს შესაბამისი მასალა.

ჩვენი მიმდინარე ამოცანები

მიუხედავად იმისა, რომ ქართული ენის მოდელირების თვალსაზრისით, ასოციაცია კვლავ არაერთი გამოწვევის წინაშე დგას, 2013 წელი ჩვენთვის საეტაპოა, რამდენადაც ენის მოდელირების პარალელურად, თვისებრივად ახალი ამოცანის განხორციელებას შევუდექით. ეს არის **ქართულ-ინგლისური სიტყვასიტყვითი თარგმნითი ლექსიკონის შექმნა**. 2013 წელს ასოციაციას შემოუერთდა პროფესორი მანანა რუსიშვილი. იგი გახლავთ თბილისის სახელმწიფო უნივერსიტეტის ინგლისური ფილოლოგიის დეპარტამენტის ხელმძღვანელი, დასავლეთ ევროპის ენებისა და ლიტერატურის-სასწავლო-სამეცნიერო ინსტიტუტის ხელმძღვანელი და საქართველოს ინგლისური ენის მასწავლებელთა ასოციაციის პრეზიდენტი. ვფიქრობთ, რომ ენის მოდელირების ასოციაციის შექმნილი ქართული ენის მოდელის, ქალბატონი მანანას ცოდნისა და გამოცდილების წყალობით და რუსთაველის ფონდის ხელშეწყობით, აღნიშნულ

ამოცანას წარმატებით გადავწყვეტ. იგეგმება 3 ათასამდე ზმნური და 25,000-მდე არაზმნური ლექსიკური ერთეულის დამუშავება. ლექსიკონის ზომა ქართული ლექსიკოგრაფიისთვის უპრეცედენტოდ დიდი იქნება - იგი 30 მილიონამდე ქართულ სიტყვაფორმას შეუსაბამებს ინგლისურ შესატყვისს. ამასთანავე ეს იქნება მნიშვნელოვანი წინგადადგმული ნაბიჯი მანქანური თარგამის პროგრამის შექმნის გზაზე.

აღსანიშნავია, რომ „ენის მოდელირების ასოციაცია“ აქტიურად თანამშრომლობს Google-თან. 2012 წელს შეხვედრა შედგა Google-ის წარმომადგენელთან. შეხვედრის ძირითადი თემა იყო Google Translate-ის ქართული მოდულის სრულყოფის საკითხი. აღნიშნული მოდული გარკვეული ხარვეზებით მუშაობს, რამდენადაც იგი მხოლოდ სტატისტიკურ მეთოდებზე დაყრდნობითაა შექმნილი. Google Translate-ის სრულყოფის მიზნით, ასოციაციამ Google-ს გადასცა ჩვენს ხელთ არსებული ქართულ-ინგლისური ლექსიკური ერთეულების შესატყვისების ელექტრონული ბაზა (91,576 ჩანაწერი).

დასკვნის სახით გვინდა აღვნიშნოთ, რომ მანქანური თარგმანის პროგრამის შექმნა უცხოელი სპეციალისტების აქტიური ჩართულობის გარეშე ყოვლად წარმოუდგენელია. მსოფლიომ უდიდესი გამოცდილება დააგროვა ამ მიმართულებით და ჩვენ მაქსიმალურად უნდა გამოვიყენოთ ეს გამოცდილება. ქართულ-ინგლისურ სიტყვათისტყვითი ლექსიკონის შექმნა არის ის მაქსიმუმი, რაც ადგილობრივი ძალებით, სახელმწიფო გრანტების არსებული ფორმატით შეიძლება შეიქმნას. მსგავსი ფორმატით პრაქტიკულად წარმოუდგენელია მაღალი კვალიფიკაციის უცხოელ ექსპერტების ჩართულობა. მეტსაც ვიტყვით: მანქანური თარგმანის ამოცანის გადაწყვეტა მოითხოვს უმაღლესი კვალიფიკაციის ენათმეცნიერების, მათემატიკოსების და პროგრამისტების ჩართულობას, რაც არსებული საგრანტო სისტემის შეზღუდვების პირობებში, შეუძლებელია. შეუძლებელია ქვეყანა ამ მასშტაბის და მნიშვნელობის ამოცანის გადასაწყვეტად ენთუზიასტების იმედზე იყოს. ამ საკითხის მიმართ აუცილებელია ერთიანი სახელმწიფო ხედვის ჩამოყალიბება და სისტემური, კოორდინირებული მუშაობა.

აუცილებელია კიდევ ერთხელ გავიაზროთ, თუ რა საფრთხის წინაშე დგას ქართული ენა და რას უქადის ეს საფრთხე ჩვენი ერის მომავალს. მანქანური თარგმანის პროგრამა ვერ შეიქმნება რამოდენიმე წელიწადში. პრაქტიკულად გამოყენებადი პროგრამის შექმნას შესაძლოა ათწლეულებიც კი დასჭირდეს. ამ ათწლეულებმა კი, შესაძლოა, ქართულ ენას განაჩენი გამოუტანოს. დრმად გვწამს, რომ სახელმწიფო ისევე უნდა ზრუნავდეს ენის შენარჩუნებაზე, როგორც ქვეყნის უსაფრთხოებაზე. იმედი გვაქვს, რომ მანქანური თარგმანის პროგრამის შექმნას საქართველოს ხელისუფლება სტრატეგიულ ამოცანათა რიცხვს მიაკუთვნებს.