

ტარიელ ფუტყარაძე

ქართული ენის ტექნოლოგიები: პირველი რიგის გადაუდებელი ამოცანები¹

თქვენო უწმინდესობავ, მეუფენო, ქალბატონებო და ბატონებო, პირველ რიგში, გილოცავთ 14 აპრილს - საქართველოს სახელმწიფო ენის დღეს.

1978 წლის 14 აპრილს, საქართველოს მოსახლეობის ღირსეულმა შვილებმა, ქართველებმაც და სხვა ეთნიკური ჯგუფების წარმომადგენლებმაც, დაიცვეს საქართველოს სახელმწიფო ენის სტატუსი, შესაბამისად, **14 აპრილი არის საქართველოს სახელმწიფო ენის დაცვის დღე** და არა მხოლოდ ქართველთა დედაენის დღე.

14 აპრილი დღემდე ოფიციალურად კვალიფიცირდება, როგორც ქართველთა დედაენის დღე; ასეთი სახელდება ნაკლებად იძლევა საშუალებას, 14 აპრილი იზეიმონ საქართველოს მოქალაქე ეთნიკურმა სომხებმა, აზერბაიჯანლებმა, რუსებმა, ოსებმა, აფხაზებმა და ა.შ. **დედაენად მიჩნევის შემთხვევაში** ამ დღეს რა უნდა იზეიმონ, მაგალითად, საქართველოს მოქალაქე სომხებმა?

თუკი 14 აპრილი ხელისუფლების მიერ ოფიციალურად გამოცხადდება საქართველოს სახელმწიფო ენის დღედ, დარწმუნებული ვარ, საქართველოს მოქალაქენი ამ დღეს **პატივისცემით აღნიშნავენ** არა მხოლოდ თბილისში, ბათუმში, ზუგდიდსა თუ თელავში, არამედ, ახალქალაქშიც, მარნეულშიც, წალკაშიც... მოვა დრო, ცხინვალშიც, გალშიც, სოხუმშიც და ა.შ.

ვფიქრობ, 14 აპრილის სახელდება მხოლოდ ქართველთა დედაენის დღედ აკნინებს 35 წლის წინანდელი მოვლენის **ზოგადსახელმწიფოებრივ მნიშვნელობას** და ხელს არ უწყობს ჩვენი ეთნიკური უმცირესობების ინტეგრაციას ქართულენოვან სახელმწიფოში.

აქვე აღვნიშნავ:

მრავალი საუკუნეა, **ქართველები დედაენის დღეს ვუწოდებთ ლაზარეს შაბათს**; წელს ლაზარეს შაბათი - ქართული ენის დღე - არის 27 აპრილს.

ბოლო პერიოდში ხელისუფლების მიერ არაადეკვატურად ნაკლები ყურადღება ეთმობა საქართველოს სახელმწიფო ენას:

¹ 2013 წლის 14 აპრილს საქართველოს საპატრიარქოსთან არსებული ახალგაზრდობის სულიერი და ინტელექტუალური განვითარების ცენტრში (სამების ახალგაზრდული ცენტრი) წაკითხული ტექსტი /მიმართვა ეკლესიისა და ხელისუფლების პირველ პირებს და მოხსენება ქართული ენის ტექნოლოგიების შესახებ/.

დედა ენა არის ნაციის არსებობის საფუძველი, ხოლო სახელმწიფო ენა ქვეყნის ღირსების საზომია; შესაბამისად, ყველა ნორმალური ხელისუფლება განსაკუთრებულ ყურადღებას აქცევს **სახელმწიფოს მთავარ ენას**, პარალელურად, უცხო ენებსა თუ უმცირესობების ენებს; სამწუხაროდ, დღეს საქართველოში ბევრს ლაპარაკობენ რეგიონული თუ უცხო ენების შესახებ; მაგალითად, საქართველოს უსაფრთხოების ოფიციალურ კონცეფციაში გათვალისწინებულია ინგლისური ენის სწავლების აუცილებლობა, მაგრამ გააზრებული არ არის საქართველოს სახელმწიფო ენის სტატუსის პერსპექტივა; კერძოდ, ქართული ენა, როგორც სახელმწიფო ენა **დე ფაქტო** არ ფუნქციონირებს საქართველოს ტერიტორიის მესამედზე, დანარჩენ ტერიტორიაზე კი მისი დაცულობის ხარისხი კრიტიკას ვერ უძლებს; უსაფრთხოების კონცეფციაში არაფერია ნათქვამი ამ მიმართულებით.

ახლა უფრო კონკრეტულად ჩემი დღევანდელი მოხსენების თემის შესახებ:

დღეს ქართული ენა სულ ახალი გამოწვევის წინაშე დგას: თანამედროვე ტექნოლოგიები რადიკალურად შეცვლის მსოფლიოს ენათა მომავალს.

საქართველოში ჯერჯერობით არ არსებობს ქართული ენის სრულმასშტაბიან ტექნოლოგიზებაზე ორიენტირებული არც ერთი ინსტიტუტი; შედეგად, ქართველი ერი და საქართველოს სახელმწიფო ყველაზე ნაკლებადაა მომზადებული ამ დიდი გამოწვევისთვის.

თანამედროვე თუ მომავალი ტექნოლოგიური ნახტომების პირობებში ქართული სალიტერატურო ენა კულტურულ ენათა რიგში რომ დარჩეს, საამისოდ აუცილებელია, დროულად განხორციელდეს ჯერ ნაწერი, შემდეგ კი, უახლოეს მომავალში, ზეპირი ტექსტის ავტომატური თარგმანი; ეს შესაძლებელს გახდის, რომ საქართველოს მოქალაქემ თავის სახელმწიფო ენაზე მიიღოს გლობალურ ქსელში არსებული ინფორმაციები. სახელმწიფო ენა ქვეყნის ღირსების საზომია, თანამედროვე სახელმწიფოს ღირსების ერთ-ერთი უმთავრესი საზომია სახელმწიფო ენის ტექნოლოგიზების დონე; გარდა ამისა, მხოლოდ ქართული ენის თანამედროვე ტექნოლოგიების ენად ქცევის შემთხვევაში შეინარჩუნებს ქართველი ერი თავისი კულტურის განმსაზღვრელ უმთავრეს ფენომენს - დედაენას; შესაბამისად, ავტომატური თარგმანის პარალელურად, აუცილებელია მომზადდეს კომპიუტერული პროგრამები, რომლებიც უზრუნველყოფს ქართული ენის სრულმასშტაბიან ტექნოლოგიზებას.

თავის დროზე მსოფლიოს მრავალი ენა დაიკარგა იმიტომ, რომ მისმა მფლობელმა ხალხმა დროულად ვერ შექმნა დამწერლობა

თავისი დედაენისათვის; თანამედროვე პირობებში ბევრი მრავალსაუკუნოვანი სამწიგნობრო ტრადიციის ენა და სახელმწიფო ენა გაქრება, თუკი მოცემული ენა დროულად ვერ გახდება ინტერნეტისივრცის, ავტომატური თარგმანების, ულტრათანამედროვე ტექნიკის ზეპირი ბრძანებით მართვის ენა; კერძოდ, ვფიქრობ, აუცილებელია, სახელმწიფოს პროგრამული დაფინანსებით (და არა, მოკლევადიანი საგრანტო წესით) დროულად განხორციელდეს მრავალპროფილიანი გრძელვადიანი პროექტი, რომლის პროდუქცია იქნება კომპიუტერული პროგრამები, რომლებიც:

- ა. ქართულ ზეპირ მეტყველებას გადააქცევს ნაბეჭდ ტექსტად,
- ბ. ავტომატურად ამეტყველებს ქართულ ნაბეჭდ ელექტრონულ ტექსტს,
- გ. გააციფრებს ნაბეჭდ პროდუქციასა და ხელნაწერებს,
- დ. განახორციელებს სრულყოფილ ავტომატურ ქართულენოვან მართლწერას,
- ე. განახორციელებს ავტომატურ თარგმანს ქართულიდან სხვა ენებზე და პირიქით...

წარმოდგენილი ამოცანების განსახორციელებლად ასევე პროგრამული დაფინანსებით სასწრაფოდაა შესაქმნელი:

- ქართული ენის დიდი ელექტრონული განმარტებითი ლექსიკონი /სადაც შევა როგორც სალიტერატურო ენის, ასევე, ქართველთა დიალექტების ლექსიკაც/;
- ელექტრონული მრავალდარგოვანი ტერმინოლოგიური ლექსიკონი;
- სამწიგნობრო ენის ნორმების ინტერაქტიული ვებპორტალი;
- საკუთარი სახელების ნორმატიული ელექტრონული ლექსიკონი;
- ქართული ენის ტექსტების ეროვნული კორპუსი, სადაც შევა ქართული სალიტერატურო ენის მასალაც და ქართველთა დიალექტური მონაცემებიც /მეგრულ-ლაზური და სვანური იდიომების ჩათვლით² და სხვ.

² ქართული ტექსტების ეროვნული კორპუსი (ქართველური დიალექტური მონაცემებით) ქართულ ენას შეუნარჩუნებს ისტორიულ ეროვნულ (გამაერთიანებლურ) და კულტურულ ფუნქციას.

დღეს არის მცდელობა, ქართული ტექსტების კორპუსის პარალელურად დაიგეგმოს /და შეიქმნას/ მეგრული, სვანური და ლაზური ტექსტების დამოუკიდებელი კორპუსებიც. უახლოეს პერიოდში ენობრივი კორპუსი იქცევა ერის /ეთნიკური ჯგუფების/ იდენტობის უმთავრეს განმსაზღვრელ ფაქტორად, შესაბამისად, ქართული ტექსტების ეროვნული კორპუსის ნაცვლად ოთხი დამოუკიდებელი “ენის” კორპუსის (ქართულის, მეგრულის, ლაზურის, სვანურის) შექმნის მცდელობა მოემსახურება მავანთა ინტერესს,

საქართველოში ქართული ენა არ არის მხოლოდ ერთ-ერთი სასწავლო საგანი; იგი სხვა საგანთა სწავლების ენაცაა; დავიმოწმებ პროფესორ ნათელა მაღლაკელიძისა და პროფესორ ნანა ლოლაძის კვლევების შედეგებს და მოკლედ ვიტყვი:

არსებული ეროვნული სასწავლო გეგმა სერიოზულ ცვლილებებს მოითხოვს. საჭიროა შეიქმნას ქართული სალიტერატურო ენის სწავლების ერთიანი კონცეფცია სწავლების ყველა საფეხურისათვის; ამასთან, საგანგებო ღონისძიებებია გასატარებელი მასწავლებელთა მაღალ დონეზე მომზადებისათვის და არა მხოლოდ - გადამზადებისათვის.

ასევე, აუცილებელია, შემუშავდეს სახელმწიფოს პრაგმატული მიდგომა ქართველთა ზეპირი საშინაო მეტყველებების გადასარჩენად, რომელთაც ნამდვილად ემუქრებათ საფრთხე; კერძოდ, განსაკუთრებით დიდი საფრთხის წინაშე არის: სვანური, ტაოური, ლივანური, იმერხეული, რაჭული, ხევსურული, თუშური, მოხეური... ქართველთა ეს კილოები ჩვენს თვალწინ ქრებიან! შესაბამისად, აუცილებლად მიგვაჩნია, სასკოლო საგნად შემოვიღოთ: **“ქართველთა ენობრივი სამყარო”**, რომლის პირველ ნაწილში წარმოდგენილი იქნება ყველა ქართველური დიალექტის გავრცელების არეალი და მეტყველების ნიმუშები, ხოლო მეორე ნაწილში შევა სპეციალური კითხვარი და ინსტრუქცია, რომ მოცემული ქართველური კილოს მფლობელს საშუალება ჰქონდეს, შეისწავლოს და წარმოაჩინოს მისი კუთხური მეტყველების ლექსიკა.

ენობრივ-ეთნიკურ-კულტურულად დაანაწევროს მრავალსაუკუნოვანი ქართველი ერი. . ასევე, პრობლემურია არნ. ჩიქობავას ენათმეცნიერების ინსტიტუტის პროექტიც: “საქართველოს ლინგვისტური პორტრეტი”, რომლის მიხედვითაც დამოუკიდებელ ენებად ითვლება მეგრულ-ლაზური და სვანური; აქ სხვა წინააღმდეგობაცაა: საქართველოს მოსახლეობის “ლინგვისტური პორტრეტი” ვერ შეიქმნება მხოლოდ ქართველური მასალით.

ვფიქრობ, ლოგიკურია, მეგრულ-სვანური-ლაზური დიალექტური მასალა, სხვა ქართველურ დიალექტურ (ხევსურულ, იმერულ, ტაოურ, ჰერულ, კახურ, მესხურ...) მონაცემებთან ერთად შევიდეს ქართული ენის ერთიან ეროვნული კორპუსში **დიალექტების სტატუსით**; კერძოდ:

სხვა მრავალი ენის მსგავსად, ქართველურ ენობრივ სამყაროშიც ენა-კილოების სტატუსი ჩვენ მიერ სოციოლინგვისტური კრიტერიუმებით ფასდება, შესაბამისად, ჩვენი პროექტის მიხედვით **ქართული ტექსტების ეროვნულ კორპუსს** აქვს ქართული სტანდარტული ენისა და ქართველური დიალექტური მასალის განყოფილებები, რომელთა ლექსიკური ფონდები ქართველთა სამწიგნობრო ენის ერთიან პერსპექტიულ სალექსიკონო ბაზადაა გააზრებული.

იმედით აღვნიშნავ, რომ საქართველოს კათოლიკოს-პატრიარქის ლოცვა-კურთხევითა მომზადებულ დღევანდელ სამეცნიერო კონფერენციაზე **სპეციალისტთა მიერ წარმოდგენილი დასკვნები** აკადემიურ საფუძველს მოუმზადებს ახალ ხელისუფლებას, სისტემურად დაგეგმოს და გადადგას ის აუცილებელი ნაბიჯები, რომლებიც უზრუნველყოფს:

1. თანამედროვე პირობებში ქართული ენის კულტურული ფუნქციის შენარჩუნებას,
2. ქართული ენის, როგორც სახელმწიფო ენის უფლებების დაცვას, სწავლებასა და კვლევას,
3. ქართველთა სამწიგნობრო ენის მასაზრდოებელი ნაკადულების - ქართველთა დიალექტების - გადარჩენასა და კვლევას.

კონკრეტული ამოცანები ასე შეიძლება ჩამოყალიბდეს:

1. ქართული ელექტრონული ნაბეჭდი ტექსტის ამეტყველება:
 - 1.1. სინტეზატორი ალოფონური ბაზის საფუძველზე;
 - 1.2. სინტეზატორი მარცვლოვანი ბაზის საფუძველზე;
 - 1.3. სინტეზატორი დიფონური ბაზის საფუძველზე და სხვ.
2. ქართული მეტყველების ნაბეჭ ტექსტად ქცევა:
 - 2.1. კარნახით ტექსტის ჩაწერა
 - 2.2. ტექნიკის ხმით მართვა და სხვ.
3. უკვე დაბეჭდილი წიგნების/თუ გაზეთების/ ტექსტებისა და ხელნაწერების ელექტრონულ ნაწერ ტექსტად ქცევა (გაციფრება);
4. ქართული ენის მორფოლოგიურ და სინტაქსურ პროცესორებზე დაფუძნებული მართლმწერის მომზადება;
5. ქართულ-ინგლისური ორენოვანი **Transfer** ტიპის თარგმანის განხორციელება /ნაწერი ტექსტისა/;
 - 5.1. კონკრეტულ ლოგიკურ გრამატიკულ მოდელზე ორიენტირებული ქართული ენის მორფოლოგიური პროცესორის აგება;
 - 5.2. კონკრეტულ გრამატიკულ მოდელზე არაორიენტირებული ქართული ენის მორფოლოგიური პროცესორის აგება;
 - 5.3. ქართული ენის სინტაქსური პროცესორის აგება;
6. ქართულ-ინგლისური ორენოვანი თარგმანის განხორციელება ე.წ. ლოგიკურ-მათემატიკური მეტა-ენის შუამავლობით;
7. ქართული ენის ლექსიკურ მონაცემთა ელექტრონული ბაზის შექმნა:
 - 7.1. ქართული ენის განმარტებითი ლექსიკონი, ქართული ენის კომბინირებული განმარტებითი ლექსიკონი.../

7.2. ქართულ-ინგლისური და სხვა ორენოვანი ელექტრონული ლექსიკონების მოპოვება /შექმნა/

7.3. უცხოური ტერმინების თარგმნის მუდმივი ცენტრის ამოქმედება

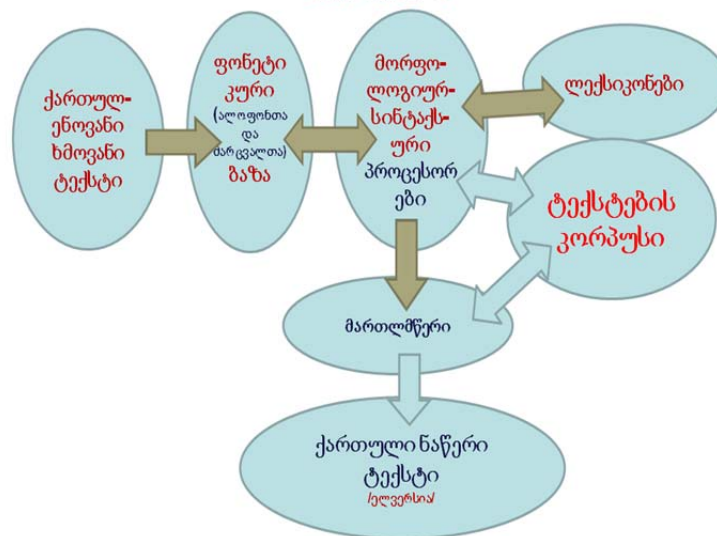
8. ქართული ენის რეპრეზენტაციული (მომცველობითი) და დაბალანსებული (პრესა, ბელეტრისტიკა, საქმიანი დოკუმენტაცია, გვარსახელები, რეესტრის სხვა მონაცემები და სხვ.) ტექსტების კორპუსის შექმნა:

8.1. მორფოლოგიურ და სინტაქსურ პროცესორებზე ორიენტირებული ტექსტების კორპუსი /პროცესორი - საძიებო სისტემა, რომელიც მოძებნის ნებისმიერ მორფოლოგიურ ფორმანტს/კატეგორიას და სინტაქსურ სტრუქტურას/, დაადგენს სიტყვის სინშირეს: უფრო რომელ ლექსიკურ ერთეულებთან (colocation) თუ მეტყველების ნაწილებთან გვხდება (colligation); გააკეთებს რეზიუმირებას;

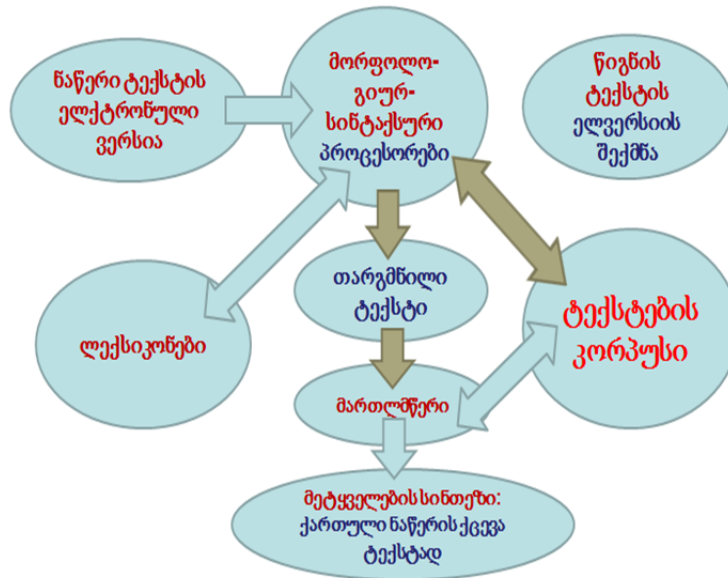
8.2. ტექსტების მორფოსინტაქსურად ანოტირებული /ან ნახევრად ანოტირებული/კორპუსი: ნახევრად ავტომატური ანალიზატორის მეშვეობით კორპუსის ანოტირება და ანოტატორთა მიერ ხელით ჩასწორება (/შემოწმება რამდენად სწორად არის ანოტირებული ამ ანალიზატორის მიერ)/...

წარმოდგენილი ამოცანები სქემატურად ასე შეიძლება გამოვსახოთ:

ქართულ ენის ტექნოლოგიზების ერთი მიმართულება -
ხმის ქცევა ტექსტად:



**ქართულ ენის ტექნოლოგიზების მეორე მიმართულება -
ელექტრონული ტექსტის თარგმნა და მეტყველების სინთეზი**



დასკვნის ნაცვლად:

საქართველოს ხელისუფლების მიერ შემუშავებულ “ეროვნული უსაფრთხოების კონცეფციის” დოკუმენტში (რომელიც დევს საქართველოს საგარეო საქმეთა სამინისტროს ვებგვერდზე³, არ ჩანს, ჩემი აზრით, **ხუთი არსებითი მიმართულება**:

ტ. საქართველო ენობრივ-ეთნიკური ერთიანობის შენარჩუნებისათვის ზრუნვა;

ჟ. ქართული ენის, როგორც სახელმწიფო ენის, ფუნქციონირების აღდგენის სტრატეგიის შემუშავება და უმცირესობებისათვის საქართველოს სახელმწიფო ენის სწავლებით ინტეგრაციის რეალური პირობების შექმნა;

ღ. ქართული ენის ტექნოლოგიზება; ამ კუთხით საქართველოს განათლების სისტემის დაგეგმვა საქართველოს სახელმწიფო ინტერესების გათვალისწინებით;

³ http://www.mfa.gov.ge/index.php?lang_id=GEO&sec_id=12 (2013 წლის მონაცემებით).

ბ. საქართველოს სახელმწიფოობრიობის ისტორიისთვის ქართულენოვანი მართლმადიდებლობისა და საქართველოს ეკლესიის განსაკუთრებული როლის აღიარება;

წ. სამხრეთ კავკასიაში (ტაო-კლარჯეთი, ლაზეთი) მცხოვრები ავტოქთონი ქართველების ეთნიკური იდენტობისა და ქართველური კულტურული მემკვიდრეობისათვის ზრუნვა.

კაცობრიობის ისტორია ხალხების **გაჩენისა და გაქრობის** მძიმე ქრონიკაა. ამ ფონზე სასწაულს ჰგავს ქართველი ერის 33-საუკუნოვანი სიცოცხლე, რომელიც, პირველ რიგში, ქართული ენითაა განხორციელებული... უფლის შეწევნით, დიდი კულტურული რევოლუციების ჟამს დროული რეაგირებით ხდებოდა ქართული ენისა და კულტურის ასვლა ახალ სიმაღლეზე - ჩვენს მიერ ნახსენები **სასწაული ქართველების ღირსეული წარმომადგენლების ხელით ხორციელდებოდა; დღესაც** დიდი კულტურული ნახტომი ხორციელდება ენების ტექნოლოგიზებით - მხოლოდ ინტერნეტსივრცის ენად ქცევა გადაარჩენს ქართულ ენას, შესაბამისად, ქართველ ერს.

იმედია, **უახლოეს მომავალში განხორციელდება ქართული ენის ტექნოლოგიზება;** შესაბამისად, კიდევ დიდხანს იქნება ქართული ენა ქართველთა ეროვნული სიცოცხლე და საქართველოს სახელმწიფოს უმთავრესი ღირსება!

Tariel Putkaradze

Technologizing of Georgian Language/the Most Urgent Tasks/

S u m m a r y

Modern technology radically changes the prospects of the world languages:

Numerous languages in the world, the alphabet of which had not been created, have disappeared. Even today a lot of century-old traditional and state literary languages will turn into Papuan languages if the given language is not transformed into the language of modern internet space, automatic translation, international information sources, the computer/or other machines/, technique control/oral voice orders/, etc.

There is no institution oriented on full-scale technologizing of the Georgian language so far in Georgia; consequently, the Georgian nation and the Georgian state is the least prepared for this big challenge.

In order to leave the Georgian literary language among the world's cultural languages in the condition of the contemporary or future technological leaps, it is necessary to put the automatic translation of the written and later oral texts into practice; this will enable Georgian citizens to get the information presented in the global network in their native tongue. State language is the standard for measuring the dignity of the nation, the level of technologizing of the state language is the standard of measurement of any modern state; besides, only turning Georgian into the language of modern technologies will preserve mother tongue- the defining phenomenon of the Georgian national identity; therefore, together with the automatized translation, it is necessary to prepare computer programs which will ensure the full-scale technologizing of the Georgian language; in particular, we can name the following specific tasks:

1. Turning Georgian electronic printed texts into speech:
 - 1.1. Synthesizer based on allophones;
 - 1.2. Synthesizer based on syllables;
 - 1.3. Synthesizer based on diaphones, etc.
2. Turning Georgian speech into printed texts:
 - 2.1. Recording dictated texts;
 - 2.2. Controlling texts with voice, etc.
3. Creating digital versions of already printed books/or newspapers/ and manuscripts;

4. Preparing orthography based on the morphological and syntactical processes of the Georgian language;
 5. Preparing Georgian-English bilingual Transfer-type translation /written texts/;
 - 5.1. Constructing Georgian language morphological processor oriented on concrete logical grammatical models;
 - 5.2. Constructing Georgian language morphological processor which is not oriented on concrete logical grammatical models;
 - 5.3. Constructing Georgian language syntactical processor;
 6. Conducting Georgian-English bilingual translation with the help of the so-called logical-mathematical meta-language;
 7. Creating electronic database of Georgian vocabulary;
 - 7.1. Georgian language explanatory dictionary, Georgian language combined explanatory dictionary...
 - 7.2. Creating Georgian-English and other bilingual electronic dictionaries;
 - 7.3. Enactment of a permanent center which will work on translation of foreign terms;
 8. Creating a corpus of comprehensive and balanced texts (press, fiction, business documents, names, registry and other data);
 - 8.1. Corpus of texts oriented on morphological and syntactical processors/processor – a search engine which will find any morphological formant/category and syntactical structure/, will determine the frequency of words: collocation or colligation; will create summaries;
 - 8.2. Morphologically annotated/partly-annotated/ corpora of texts: correction will be made with the help of semi-automated analyzers (they will check whether this is properly annotated by this analyzer).
- The mankind history is a severe chronicle of **emergence** and **disappearance** of peoples. Regarding the background, the 33-century existence of the Georgian nation, mainly thanks to the Georgian language, is like a miracle. With the help of God, Georgian language and culture managed to climb to new heights thanks to the timely reactions to major cultural revolutions. The abovementioned miracles were carried out by dignified Georgians; a great cultural leap is made today by technologizing languages and turning Georgian into the language of internet that will enable it to survive; consequently, the Georgian language will be the main national dignity of the Georgian nation!