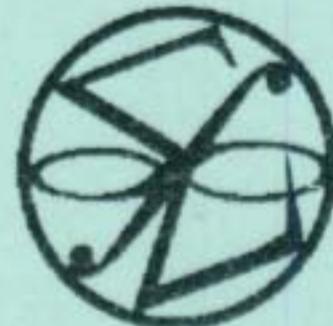


III კონფერენცია

ბუნებრივ ენათა დამუშავება

ქართული ენა და კომპიუტერული

ტექნოლოგიები



Conference III

Natural Language Processing

The Georgian Language and Computer Technologies

საქართველოს მეცნიერებათა აკადემიის არნ. ჩიქობავას
სახელობის ენათმეცნიერების ინსტიტუტი

**ბუნებრივ ენათა დამუშავება
ქართული ენა და კომპიუტერული
ტექნოლოგიები**

2005 წლის 29 ივნისი — 1 ივლისი

კონფერენციის მასალები



„ქართული ენა“
თბილისი
2005

მუშაობის გეგმა
თთხშაბათი, 29 ივნისი

- 11.00 გ. კვარაცხელია — შესაფალი სიტყვა
11.30 თ. უთურგაიძე, ლ. ეზუგბაია, თ. ვაშაკიძე, მ. მანქვალაძე —
ტრანსლაციური ბანკისა და დისტრიბუციული აპარატისათვის.
12.00 მ. კვიმსაძე, თ. ქავთარაძე, მ. ტაბიძე — სასაქონლო ნიშნების სის-
ტემატიზაციისა და მსგავსებათა ძიების საკითხისათვის.
12.30 მ. ჩაჩანიძე, მ. ცისკარიშვილი — წოვა-თუშური (ბაქტური) ენის
მონაცემთა ელექტრონული ბაზის შექმნის ძირითადი ასპექტები
- 13.00-14.00 — შესვენება
- 14.00 ი. მოდებაძე, თ. ციციშვილი — ქართველოლოგიურ ლიტერატურათ-
მცოდნეობით მონაცემთა ბაზის დამუშავების თავისებურებანი და
პრობლემები
- 14.30 ქ. გოჩიტაშვილი, გ. შაბაშვილი — WC-ში განთავსებული ქართული
ტექსტების ელექტრონული დამუშავების პრინციპებისათვის
- 15.00 ლ. გარგველანი — ენის კომპიუტერული პროცესორების სემანტიკური
ეტაპის შესახებ
- 15.30-16.00 დისკუსია

სუთშაბათი, 30 ივნისი

- 11.00 **გ. ანთიძე, დ. მიშელაშვილი** — პროგრამული უზრუნველყოფა ბუნებრივ ენაზე ჩაწერილი ტექსტების მორფოლოგიური და სინტაქსური ანალიზისათვის
- 11.30 **ქ. ფხაკაძე, ვ. ლეგიაშვილი, რ. სხირტლაძე, გ. კანდელაკი, ვ. ფხაკაძე** — ბუნებრივი ქართული ენობრივი სისტემის პირდაპირი ფორმალურ-ლოგიკური აღწერა და მისი ზოგიერთი გამოყენება
- 12.00 **გ. ჩიჩუა** მეტყველების კომპიუტერული გამოცნობა
- 12.30 **ჭ. მაჩაიძე, თ. ჩიჩუა, ქ. ბიწაძე, ლ. ალთუნაშვილი, ნ. პავლიაშვილი, რ. კვირკველაძა** — დისტანციური განათლების პროგრამული გარსების გაქართულების პრობლემების შესახებ (მოდულური ობიექტ-ორიენტირებული დისტანციური სწავლების გარემოს (MOODLE) მაგალითზე)
- 13.00-14.00 — **შესვენება**
- 14.00 **გ. კობაიძე** — ქართული ენის და, კურძოდ, ქართული ზმნის არაქართულენოვანი მოსწავლეებისთვის სწავლების შესახებ
- 14.30 **გ. ბერიძე, დ. ნადარაძა** — მონაცემთა ბაზა დიალექტური ტექსტების კორპუსის ელექტრონული ვერსიისათვის
- 15.00 **გ. ბერიძე, ლ. ბაკურაძე, ლ. გოგოხია, ი. უგულავა** — მონაცემთა ბაზა დილექტური ლექსიკის კარტოგრაფირებისათვის.
- 15.30 **ნ. ხუციშვილი** — „ბრალთგადამხდელობითი“ ნაწილი XI-XV საუკუნეების ქართულ საისტორიო საბუთებში (კომპიუტერული ანალიზი)
- 16.00-16.30 — **დისკუსია**

პარასკევი, 1 ოცნები

11.00 გ. შერგაშიძე — ქართული ენა და საინფორმაციო და საკომუნიკაციო ტექნიკური გენერატორი

11.30 ე. ჭერენაია, თ. ჩხენკელი, გ. აფრიდონიძე, გ. ვარაშიშვილი — ელექტრონული ბიბლიოთეკა საქართველოს პარლამენტის ილია ჭავჭავაძის სახელობის ეროვნულ ბიბლიოთეკაში

12.00 დ. თვალთვაძე, ს. სარჯველაძე — ძველი ქართული ენის ძეგლების დიგიტალიზაცია (ოთხთავის ქართული თარგმანის კრიტიკული ტექსტის დამდგენი ჯგუფის მუშაობის შედეგები)

12.30 მ. კაპანაძე, ი. გაბუნია — სასრული მდგრამარეობის ტრანსდუსერი ქართული ენის არსებითი სახელების ანალიზისა და სინთეზისათვის

13.00-14.00 — შესვენება

14.00 ლ. ლორთქიფანიძე — რუსული ენის მორფოლოგიური წარმოდგენის უნიფიკაციის ავტომატური სისტემა

14.30 ნ. ლოლაძე, ქ. დათუკიშვილი მ. ზაკალაშვილი — ქართული ენის კორპუსის სტრუქტურა

15.00 ქ. დათუკიშვილი, ნ. ლოლაძე მ. ზაკალაშვილი — მორფოლოგიური ინფორმაციის სტრუქტურა ქართული ენის კორპუსისათვის

15.30-16.00 ღისკუსია

სახელთა ზმინისართად მოქცევისას: დღე = O („დღე შრომობს...”), მთელი ზაფხული („მთელი ზაფხული ნადირობდა”).

კონსტანტინე ფხაგაძე, ვლადიმერ ლექაიაშვილი, რატი სისტემაძე,

გიორგი კანდელაძი, ვახტანგ ფხაგაძე

ი. ვეჯვას სახ. გამოყენებითი მათემატიკის ინსტიტუტი
ლოგიკისა და ენის გაერთიანებული ქართული ჯგუფი
pkhakadz@viam.du.ge

**ბუნებრივი ქართული ენობრივი სისტემის პირდაპირი
ფორმალურ-ლოგიკური აღწერა და მისი ზოგიერთი
გამოყენება**

ნაშრომში სისტემატიზებული სახით წარმოვადგენთ ბუნებრივი ქართული ენობრივი სისტემის პირდაპირი ფორმალურ-ლოგიკური აღწერის სასაყრდენე თვალსაზრისებს, ზოგად ფორმალურ-ლოგიკურ მიღობებსა და მეთოდებს, რის შემდგომაც მოკლედ მიმოვინილავთ ქართული ენის საფუძველზე ჩამოყალიბებული მეთოდების განზოგადების შესაძლებლობას არაქართველური ტიპის ენებზე. კერძოდ, ჩვენ შევეხებით ინდოეგროპული ენებისათვის ისეთ ფუნდამენტურ საკითხს, როგორიცაა შედგენილი შემასმენლის საკითხი. ქართული და ფრაგმატიული მათემატიკური ენებისათვის დამახასიათებელი მიღობებით, წინადადებაში (I) ‘**a student is reading a book**’ შემასმენლის ანუ პრედიკატის ანუ წინადადების მაყალიბებელი მთავარი სიტყვის ფუნქციას ასრულებს ‘**is**’ (‘არის’) ორადგილიანი მიმართება, მაშინ როდესაც ‘**reading a book**’ (‘მკითხველი წიგნისა’) იმავე მიღობებით მიმღეობური აღნაგობის (წარმოების) სახელური ფრაზაა და მისი სიმრავლური მნიშვნელობაა დროის აწყვო მომენტში კულა იმ მკითხველის სიმრავლე, რომლებიც წიგნს კითხულობენ. ამ გაგებით, მიმღეობური ფრაზა ‘**reading a book**’ (‘მკითხველი წიგნისა’) ფრჩხილებიანი ჩანაწერებით ფორმდება როგორც ‘**(reading) a book**’ (‘მკითხველი წიგნისა’), რაც თავის შერიც იმას ნიშავს, რომ ოპერირება წარმოებს სახელით მიმღეობაზე და არა პირიქით. ამგვარად, ამ მიღობებით და წინადა სიმრავლურ-თეორიული თვალსაზრისებით (I) წინადადება შემდეგნაირად ითარგმნება: ‘**(a student) ∈ {(reading) a book}**’, საიდანაც გასაგები ხდება, რომ მათემატიკური შესატყვისი ‘**is**’ ორადგილიანი მიმართებისა არის ‘**∈**’ პრედიკატი.

გარდა ამისა, მონტეგიოუსეული მიღობებით, რაც ზოგად არის ტოტელესეულ მიღობებსა და, შესაბამისად, სახელების ერთადგილიან პრედიკატებად გააზრებას ეყრდნობა, (I) მათემატიკურ ენაზე ითარგმნება როგორც $\exists x \text{ student}(x) \rightarrow \exists y (\text{read}(x,y) \wedge \text{book}(y))$, რის ნაცვლადაც ჩვენი მიღობები თარგმანში იძლევა შემდეგ მათემატიკურ ჩანაწერს: $\exists x_{x \in \{\text{student}\}} \exists y_{y \in \{\text{book}\}} \text{read}(x,y)$, რაც, ბუნებრივია, უკვე არსებული კლასიკუ-

რი მონტეგიუსეული მიღებისაგან ახალი, არსებითად განსხვავებული მიღებმაა.

ამგვარად, ქართული ენიდან წამოღებულმა ფორმალურ-ლოგიკურმა მიღებულებმა გარეგნულწილად უარი გვათქმევინა სუბიექტის, კოპულასა და პრედიკატის არისტოტელესეულ სასაყრდენე კატეგორიებსა და მონტეგიუსეულ ზოგად მიღების ზე და მათ ნაცვლად ბევრად უფრო ბუნებრივი, მათგატიკური მიღები შემოგვთავაზა.

ნაშრომში თეორიული საკითხების სისტემური და სიდრმისეული მიმოხილვის გარდა, მოკლედ წარმოვაჩენთ ამ თეორიული მიღების გამოყენებით გაკათებულ ექსპრიმენტულ პროგრამულ რეალიზაციებს:

(I) ლოგიკისა და ენის გაერთიანებულ ქართულ ჯგუფში კ. ფხა-კაძის ხელმძღვანელობით შემუშავებული თეორიული მიღების საფუძველზე რ. სხიორტლაძის მიერ მომზადდა პროგრამა, რომელიც ჩვენ მიერ გაგებულია როგორც ქართული ენის ბირთვული ნაწილის ფორმალური მოდელის კომპიუტერული რეალიზაცია. ამასთან, პროგრამა იძლევა მარტივი წინადაღებისა და სახელური ფრაზების საქმაოდ ფართო კლასისათვის სინტაქსური საფეხურის მართლმწერის კონსტრუირების საშუალებას. სისტემა აგებულია ეწ. ტიპიზებული ლექსიკონის საფუძველზე. ამასთან, ჯერჯერობით რეალიზაციაშია მხოლოდ ზოგადი N, A და V ტიპები. ამასთან, სისტემაში თითოეული მათგანი იძლევა N-ა, A-ა, V(N-ა, N-ა, N-ა) სახის კერძო ტიპებს, სადაც $\alpha \in \{-\text{o}, -\text{მა}, -\text{ხ}\}$. პროგრამის სინტაქსური ინტელექტური უკრდნობა ზემოთ აღწერილი ტიპებით მოცემულ ფორმალურ წესებს, რომელთაგან ზოგი ინვერსირებადი, ზოგი კი არაინვერსირებადი წესია. ასე მაგალითად, N-ა + V(N-ა, N-ა, N-ა) = V(N-ა, N-ა) სისტემაში არსებული ინვერსირებადი წესის მაგალითია, რაც თავის მხრივ იმას ნიშნავს, რომ $(N-\text{ა}, N-\text{ა}, N-\text{ა}) + N-\text{ა} = V(N-\text{ა}, N-\text{ა})$ ფორმულაც სისტემური წესია.

განსაკუთრებით აღსანიშნავია ის, რომ აქ წარმოდგენილი მიღები იძლევა მათი თანდათანობითი განვითარებისა და დაზუსტების საშუალებას.

(II) ლოგიკისა და ენის გაერთიანებულ ქართულ ჯგუფში კ. ფხა-კაძის ხელმძღვანელობით შემუშავებული თეორიული მიღების საფუძველზე გ. კანდელაკის მიერ შზადება პროგრამა, რომლის პირობითი დასახელებაა „ქართულიდან და ქართულზე მათემატიკური ენის გავლით მრავალენოვანი მთარგმნელი პროგრამული სისტემა“.

ისევე, როგორც წინა შემთხვევაში, აქაც თეორიული მიღები იძლევა რეალიზაციების თანდათანობითი გაფართოებისა და დაზუსტების შესძლებლობას. პროგრამა უკრდნობა მათემატიკურად ტიპიზებულ მონაცემებს. წინა პროგრამისაგან განსხვავებით ამ შემთხვევაში დასამუშავებელ ტიპთა კლასი უფრო სრულყოფილია. კერძოდ, N-ა, A-ა და V(N-ა, N-ა, N-ა) ტიპების გარდა მუშავდება მ-ა, ა-ა, უ-ა, უ-მ-ა სახის კვანტორული ტიპებიც. ამასთან, სისტემაში არსებით დატვირთვას იღებს თავის თავზე ეწ. წინარე სმნური სემანტიკური ერთეულის სასაფუძვლო ლინგვისტურ-ლოგიკური კატეგორია. პროგრამა

მთარგმნელობითი ინტერესების გათვალისწინებით აკაგშირებს ქართულ ენას და მათემატიკურ ენას. პროგრამაში, გარდა აღნიშნულისა, გარეგნებული მოცულობით უკვე გათვალისწინებულია პირთაშორისი მომართებების ინტელექტუალური ანალიზი.

(III) ლოგიკისა და ენის გაერთიანებულ ქართულ ჯგუფში კ.ფხავაძის ხელმძღვანელობით შემუშავებული ოცნებისული მიღვომების საუკეთესო ვ. ფხავაძის მიერ რეალიზებული პროგრამა არის პირველი ექსპერიმენტი ქართულში ინტელექტუალური, ენობრივ-ლოგიკური გამოთვლებისა. ისვევ როგორც წინა პროგრამებში, აქაც არსებითია მონაცემთა მათემატიზებული წარმოდგენა ზოგადი ფორმალურ-ლოგიკური ტიპების სახით და, შესაბამისად, ქართული ენისა და მათემატიკური ენის შინაარსული ურთიერთმიმართებები.

სისტემაში, გარდა N-ა, A-ა, V(N-ა, N-ა, N-ა, N-ა), ჸ-ა, ა-ა, ე-ა, უ-ა ტიპებისა, გათვალისწინებულია აგრეთვე შემდგები წმინდა ლოგიკური, მათემატიკური სიმბოლოები \neg , \rightarrow , \rightarrow , \wedge , \vee , \neq , \leftrightarrow , \emptyset , \in , $=$. და t (ჭეშმარიტი), f (მცდარი), i (განუსაზღვრელი) ჭეშმარიტი მნიშვნელობები. წარმოდგენილი პროგარამული სისტემა რეალიზებულია ექსელში და მას ლოგიკური ამოცანების ამომხსნელ და შემმოწმებელ სისტემას ეუწოდებთ. იგი ჯერჯერობით ამუშავებს ორი ან ერთი წინადაღებით მოცემულ ტექსტურ, ენობრივ-ლოგიკურ ამოცანებს და პირობის შესაბამისად იძლევა დასკვნას, ან იძლევა შეტყობინებას, რომ ახალი დასკვნის გაკეთება შეუძლებელია. ხოლო იმ შემთხვევაში, თუ ენობრივად მოცემულ ლოგიკურ ამოცანას თან ერთვის შემოთავაზებული პასუხიც, მაშინ სისტემა იძლევა ამ პასუხის კორექტულობა-არაკორექტულობის შეფასებას. სისტემა, ისვევ როგორც წინა სისტემები, მუშაობს მათემატიკურ ენასთან უწყვეტ კომბინაციურ კავშირში და აქაც გათვალისწინებულია რეალიზაციების თანდათანობითი გაფართოებისა და დაზუსტების შესაძლებლობები.

გია შერვაშიძე

განათლების მართვის საინფორმაციო სისტემა
guiasher@mes.gov.ge

ქართული ენა და საინფორმაციო და საკომუნიკაციო ტექნოლოგიები

მოხსენებაში განიხილება თანამედროვე საინფორმაციო და საკომუნიკაციო ტექნოლოგიებში ქართული ენის მხარდაჭერის, კომპიუტერული ტერმინოლოგიის, ღია კოდების პროგრამული უზრუნველყოფის გაქართულების სპეციფიკის და მანქანურ თარგმანისთან დაკავშირებული საკითხები.

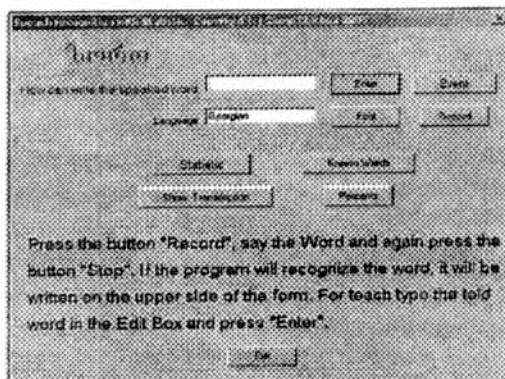
გიორგი ჩიჩუა

თბილისის სულხან-საბა ორბელიანის სახ. პედაგოგიური უნივერსიტეტი
ლოგიკისა და ენის ქართული ჯგუფი
gchichua@yahoo.com

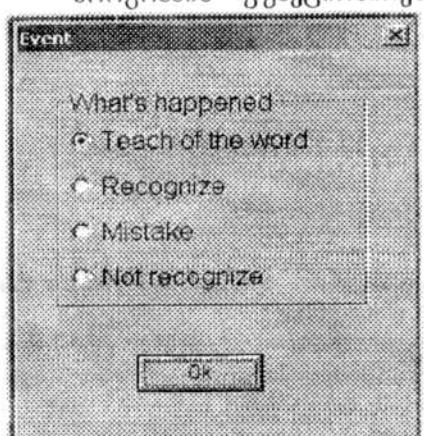
მეტყველების კომპიუტერული გამოცნობა

მეტყველების კომპიუტერული გამოცნობა არის პროცესი, რომ-
ლის შედეგად კომპიუტერი სამეტყველო სიგნალს შეუსაბამებს
ტექსტს. მეტყველების გამოცნობის არე თავისი არსით მულტიდისკიპ-
ლინარულია, მოიცავს რა მეცნიერების სხვადასხვა დარგს, როგორი-
ცაა ფიზიკა, ფიზიოლოგია, აკუსტიკა, სიგნალის დამუშავება, კომპიუ-
ტერული მეცნიერებები და სხვ. მოლაპარაკისგან დამოუკიდებელი სის-
ტემა დამუშავებულია გარკვეული ტიპის ნებისმიერ მოლაპარაკესთან
ოპერირებისათვის. ასეთი სისტემების დამუშავება განსაკუთრებით
რთულია, ამავდროულად ძვირადიორებულია და მისი სიზუსტე უფრო
დაბალია, ვიდრე მოლაპარაკეზე დამოუკიდებული სისტემებისა. მიუხე-
დავად ამისა, ისინი უფრო მოქნილია. იზოლირებულ სიტყვათა სისტე-
მა ამუშავებს ცალკეულ სიტყვებს და ამავდროულად მოითხოვს პაუ-
ზებს წარმოთქმულ სიტყვებს შორის. რეალიზების თვალსაზრისით ეს
არის გამოცნობისთვის უმარტივესი ფორმა, რადგან აღვილია სიტყვის
საწყისი და საბოლოო წერტილების პოვნა და სიტყვის წარმოთქმა არ
მოქმედებს მომდევნო სიტყვებზე. მეტყველების გამომცნობი სისტემის
ლექსიკონის ზომა მოქმედებს სისტემის კომპლექსურობაზე, პროცედუ-
რების მოთხოვნებსა და სისტემის სიზუსტეზე. სისტემის ზოგიერთი
გამოყენება მოითხოვს მხოლოდ რამდენიმესიტყვიან ლექსიკონს (მაგ.,
მხოლოდ ციფრები). სხვა სისტემები მოითხოვნ ძალიან დიდ ლექსი-
კონებს (მაგ., საკარნახო მანქანები).

განსახილები პროგრამა შექმნილია ენისა და ლოგიკის გაერ-
თიანებული ქართული ჯგუფის ფარგლებში და ახორციელებს წარმო-
თქმული ცალკეული სიტყვების გამოცნობას. პროგრამის ინტერფეისი
გამოხატულია ნახ. 1 –ზე სიტყვა “ზოროს” გამოცნობისას.

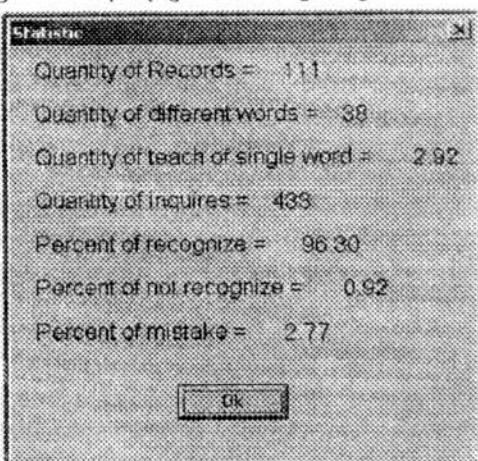


ნახ. 1

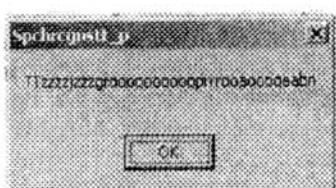


6a) b. 2

წარმოთქმის შემდეგ უნდა დაგაჭიროთ ლილაკს Event. გამოვა ნახ. 2 – ზე ნაჩვენები დიალოგური ფანჯარა, რომელშიც უნდა მივუთითოთ, რა მოხდა: სიტყვის სწავლება, გამოცნობა, შეცდომა თუ ვერგამოცნობა. შეცდომა ნიშნავს, რომ პროგრამამ დაწერა წარმოთქმული-საგან განსხვავებული სიტყვა. ხოლო ვერგამოცნობა ნიშნავს, რომ პროგრამამ არ დაწერა არცერთი სიტყვა. სტატისტიკის სანახავად უნდა დაგაჭიროთ Statistic ლილაკს. ეკრანზე გამოვა ნახ. 3-ზე ნაჩვენები ფორმა.



60k. 3



6>h 4

38-იანი გამოცდილია 38 განსხვავებული სიტყვისთვის და 433 ქასპერომენტის შეფასად მიღებული გამოცნობის პროცენტია 96.30%.

ରେବା ଶ୍ଵେତଦୟନାିରୀଳା: ଉନ୍ଦର ଡାକାଖିରିଣ
Record ଲୋଲାଙ୍କୁ, ଗ୍ରହ୍ୟତ ସିର୍ବ୍ୟବା ଡା
କାଖାଖିରିଣ Stop ଲୋଲାଙ୍କୁ. ତୁ କ୍ରମଗ-
ରାମ ଗାମରୀଚ୍ଛନ୍ଦ୍ରରେ ଫାରମତ୍ତମ୍ଭୁଲ ସି-
ତ୍ତଗାଲୁ, ଗ୍ରହ ଡାଇନ୍ଫର୍ମ୍ୟୁବା ଘୋଷମିଳି ଦ୍ୱା-
ରା ମୋର୍ଗୁବା. ତୁ କ୍ରମଗରାମାମ ଜ୍ଞାନ ଏଣ
ପିଲି ଫାରମତ୍ତମ୍ଭୁଲି ସିତ୍ତପାବ, ସାକ୍ଷି-
ରା ଗାଲ୍ଫାଗଲାନ ଗ୍ରହ. ଅନ୍ତିମତ୍ତିଲି ସି-
ତ୍ତପାବ ଉନ୍ଦର ଆଗରିବିନିତ ର୍ଯ୍ୟାନ୍ତିରିଣ୍ୟ-
ଦୀର ଗ୍ରେନ୍ଡି ଡା କାଖାଖିରିଣ Enter -ବ.
ସାଲୁରାଗଲାର ତିତର ସିତ୍ତପାବ ଗାଲ୍ଫାଗ-
ଲାନ ଏଣଜ୍ଞାନ ଅନ ସାମଜ୍ଞୟର. କ୍ରମଗରାମା
ଅନ୍ତିମତ୍ତମ୍ଭୁଲିଲାଇ ସତ୍ତାତ୍ତ୍ବିନ୍ଦିପାରି ଏନା-
ଲାନିଲି ସାମଜାଲ୍ଲାପିତ. ସତ୍ତାତ୍ତ୍ବିନ୍ଦିପାରି
ଏନାଲ୍ଲାନିଥିବେତପାବ ପ୍ରାଣିଲାଇ ସିତ୍ତପାବିଲି

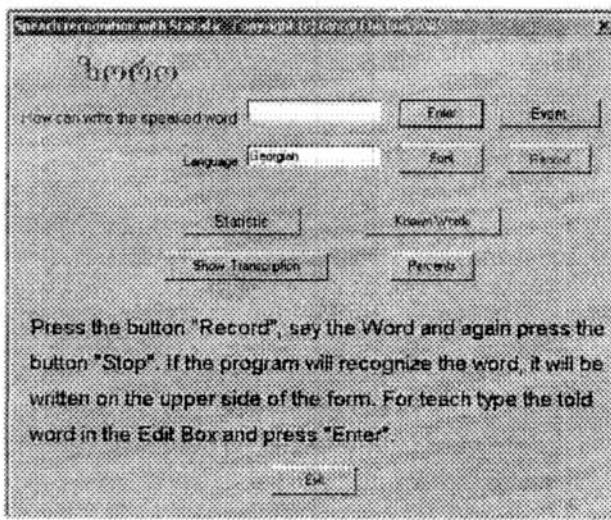
G. Chichua

Sulkhan Saba Orbeliani State Pedagogical University
Georgian Group of Logic and Language
gchichua@yahoo.com

COMPUTER SPEECH RECOGNITION

Computer speech recognition is the process by which a computer maps an acoustic speech signal to text. The field of speech recognition is inherently multi-disciplinary in nature, drawing upon various areas of study, including Physics, Physiology, Acoustics, Signal Processing and Computer Science, to name but a few. A speaker independent system is developed to operate for any speaker of a particular type. These systems are the most difficult to develop, most expensive and accuracy is lower than speaker dependent systems. However, they are more flexible. An isolated-word system operates on single words at a time - requiring a pause between saying each word. This is the simplest form of recognition to perform because the end points are easier to find and the pronunciation of a word tends not affect others. The size of vocabulary of a speech recognition system affects the complexity, processing requirements and the accuracy of the system. Some applications only require a few words (e.g. numbers only), others require very large dictionaries (e.g. dictation machines).

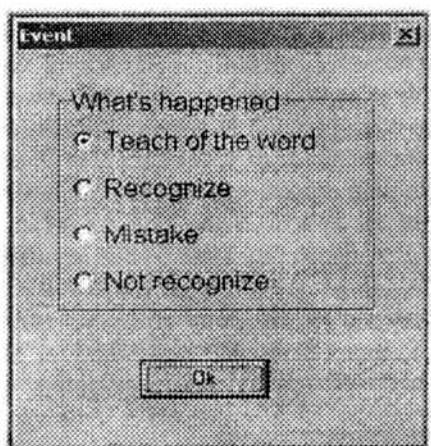
The computer program which is developed by the group of experts of the Georgian united group of Language and Logic, carries out recognition of the separately pronounced words. The interface of the program is represented on fig. 1.



Pic.1

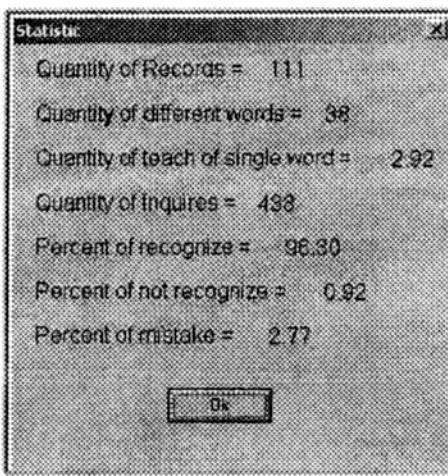
The image corresponds to the recognition of the word "Zoro", pronounced in Georgian. The program functions as follows: After clicking on the button "Record"

it is necessary to make a word, and click on the button "Stop". If the program recognizes the word, it will typed, colored in green, in the left top corner of the window, as in this event. This will occur so, if the program is trained to the pronunciation of this word. If this word is made for the first time and to this word the program is not learned yet, the program cannot distinguish it. In such occasions for the program is necessary to learn this word. For this purpose the word which had been made is necessary to type in the frame of editing and then clicking on the "Enter" button. It is desirable to teach each word two or more times. In the program is embedded the means of the data statistical processing. For carrying out of the statistical analysis after each word pronouncing must be clicked on the "Event" button.



Pic.2

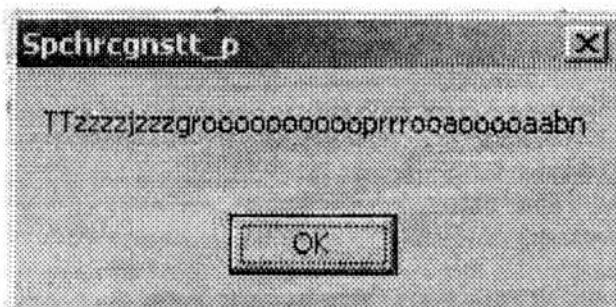
There it will be displayed the window shown on the Fig. 2, where must be specified what happened: The word teaching, recognizing, mistake or not recognizing. The "Mistake" means that the program typed word which differs from pronounced one. Also 'Not recognize' means, that no word were typed. For the checking of the statistic is required clicking on the "Statistic" button. The form shown on the Fig. 3 will be displayed.



Pic.3

Functioning of the program is based on the comparisons of the words' transcriptions. In a transcription to each letter there corresponds sequence of the several identical transcriptions. For the transcription's checking is required clicking

on the "Show Transcription" button. For example; the transcription of the word "Zoro" is shown on the Fig. 4.



Pic. 4

The program is checked up on a series of recognition of 38 various words and on the average the percentage parameter of correct recognition makes 96,3% of 433 experiments.

K. Datukishvili, N. Loladze, M. Zakalashvili
Georgian Information-Linguistic Centre
GILC@Wanex.ge

THE MORPHOLOGICAL INFORMATION STRUCTURE FOR THE GEORGIAN LANGUAGE CORPUS

Texts in the Georgian Language Corpus are annotated according to the syntactic and semantic characteristics.

The morphological information represented in the Georgian Language Corpus can have the following structure (the word-forms are referenced):

1. lemma and its morphological characteristics (morpho-phonological changes);
2. lexeme (nouns in nominative case, verbs in masdar forms);
3. morphological categories of lexemes (declination type for nouns; voice, causation and so on for masdar forms);
4. morphological categories of word-forms (case and number for nouns; person, number, tense, aspect, mood for verbs, etc.).

Such description of texts needs special inventories enabling automatic analysis and lemmatization of the texts. After such working up of the texts it is possible to gain the information according to the different morphological characteristics such as case, number, screeve and other categories.

(terminology) using key-words in all text material. We believe, the selection of these key-words represents most significant problem of this project.

K. Pkhakadze, L. Lekhiashvili, R. Skhirtladze, G. Kandelaki, V. Pkhakadze
United Georgian Group of Logic and Language

DIRECT FORMAL-LOGICAL DESCRIPTION OF THE GEORGIAN NATURAL LANGUAGE SYSTEM AND ITS SOME GENERALIZATION AND APPLICATION

In the paper we will generally and systematically represent some main points of Direct Formal-Logical Description of the Georgian Natural Language System (DFLD of GNLS) and we will try to show the possibility to apply the methods and approaches of DFLD of GNLS to the non-Georgian type languages.

According to the predicate approaches peculiar to the Georgian and Frege's type mathematical languages and according DFLD of GNLS in the sentence 'a student is reading a book' 'is' is two place predicate symbol (it is not a copula) and 'reading' is not a predicate, but the expression 'reading a book' in the interpretation '(reading) a book' is a participle phrase and gives us the set of all that readers which are in reading some book by now. Also, 'student' as noun is not one place predicate symbols, but 'a student' is a non-proper constant type noun phrase. It means that 'a student' is an element (member) of the set {student}, but it is unknown concretely who is he(she). So, according to our Georgian lingua-logical intuition and from pure set-theoretic point of view the sentence (I) can be translated as ' $(a\ student)^c \in \{(reading)\ a\ book\} \subset \{reading\}$ ' (this means that this student belong to the set of all that people which are reading a book by now) from which it is clear that the mathematical adequate of 'is' is the predicate ' \in '.

Besides, if in the sentence 'a student is reading a book' we consider '[is] reading' as two-place predicate symbol and 'a student' and 'a book' we consider as non-proper constant type noun phrases, then the mathematical translation of the sentence 'a student is reading a book' will be the formula $\exists x_{x \in \{\text{student}\}} \exists y_{y \in \{\text{book}\}} \text{read}(x, y)$ instead of the formula $(\exists x \text{student}(x) \exists y \text{read}(x, y) \wedge \text{book}(y))$.

So, the formal-logical approaches taken from the GNLS made us in some sense to decline the general Aristotelian and Montaguean interpretation of simple sentences on the basis of concepts of subject, copula and predicate.

To show applied aspects of DFLD of GNLS we shall shortly describe and demonstrate several computer programs which are constructed according that formal algorithms which has been worked out in the "United Georgian Group of Logic and Language" (supervisor Dr. K. Pkhakadze).

(1). In the program system is used formal-mathematical lingual-logical methods worked out under leadership Dr. K. Pkhakadze. The program realization belongs to R. Skhirtladze. The program has built-in dictionary of typified Georgian

words. User can add new words to it. To each word corresponds appropriate category (N with indexes for Nouns, A with indexes for Adjective, V with indexes for Verbs) according its morphology. Syntactic rules are described as algebraic operations on $N-\alpha$, $A-\alpha$, $V(N-\alpha_1 \ N-\alpha_2 \ N-\alpha_3)$ types where α is morph-markers of Georgian Language. For example:

$$N-\alpha_1 + V(N-\alpha_1 \ N-\alpha_2 \ N-\alpha_3) = V(N-\alpha_2 \ N-\alpha_3)$$

is one of the our rule which because of properties Georgian language is understood as inverse rule. This means that

$$V(N-\alpha_1 \ N-\alpha_2 \ N-\alpha_3) + N-\alpha_1 = V(N-\alpha_2 \ N-\alpha_3)$$

also is a right rule of our system.

The program can be appreciate as: (I) one of the first experimental computer model of core part of Georgian (II) One of the first experimental realization of the syntactic analyzer of the simple sentences and simple noun phrases and syntactic kind spellchecking for Georgian

(2). We have began to work on the computer program which is understood as multi-lingual translator from and into Georgian language through the mathematical language. The program is designed according that formal algorithm which are worked in the UGGLL under leadership Dr. K. Pkhakadze. The program realization belongs to G. Kandelaki. The system is mainly designed on the basis of $N-\alpha$, $A-\alpha$, $V(N-\alpha_1 \ N-\alpha_2 \ N-\alpha_3, N-\alpha_4)$, $\exists-\alpha$, $\forall-\alpha$, $\neg\exists-\alpha$, $\neg\forall-\alpha$ types. In the program system is used mathematical lingual-logical methods worked under leadership Dr. K. Pkhakadze. According this methods the concept of the pre-verbal semantic unites is one of the main category for all program system. The category gives the system possibility mechanically generate and formally described verbs by the basic types of $N-\alpha$ kinds. This gives us possibility to make syntactic synthesizer and analyzer for Georgian on the basis of $N-\alpha$, $A-\alpha$, $V(N-\alpha_1 \ N-\alpha_2 \ N-\alpha_3, N-\alpha_4)$, $\exists-\alpha$, $\forall-\alpha$, $\neg\exists-\alpha$, $\neg\forall-\alpha$ types and, also, it gives us possibility to translate Georgian phrases and sentences into mathematical language. and vice versa. This program system is also first Georgian experiment into this direction.

(3). In the program system is used general lingual-logical methods and some methods of natural logic of GNLS worked under leadership Dr. K. Pkhakadze. The program realization belongs to V. Pkhakadze. This program is realized in Excel. It solves a verbal logical examples for Georgian. This program system is working on the basis of $N-\alpha$, $A-\alpha$, $V(N-\alpha_1 \ N-\alpha_2 \ N-\alpha_3, N-\alpha_4)$, $\exists-\alpha$, $\forall-\alpha$, $\neg\exists-\alpha$, $\neg\forall-\alpha$ types, but in the program is also in used clear logical type symbols such as \neg , \wedge , \vee , \subset , \in , \cup , \cap , \exists , \forall , \emptyset , \in , $=$, t (true), f (false), i (indefinite). The program has the dictionary of the typified Georgian words. User can add any new words in the dictionary but only if it is already described type. The program is constructed so that it is assumed that the constituents of a given Georgian verbal-logical example are a well-formed Georgian expressions. According to our approaches the system first made the mathematical translation of the given Georgian verbal-logical expression and, after, according to rules of inference which are giving the system intellectual possibilities first, if it is possible, make logical inference on the mathematical level and after through retranslating given conclusion into adequate

Georgian Language expression give us the right answer of the given verbal-logical example. But, if rules of inference which are in the program systems are not able to give conclusion then program gives us information about impossibility to make conclusion. By now the system are working only verbal-logical examples with one or two premises. The system is also able to check logically a argument and if conclusion is not right and if a conclusion is inferable from the premises then program gives us right conclusion. This program system is also first Georgian experiment into this direction.

G. Shervashidze

Project – "Ilia Chavchavadze"

Education Management Information System – coordinator

guiasher@mes.gov.ge

GEORGIAN LANGUAGE AND INFORMATION AND COMMUNICATION TECHNOLOGIES

The problems of Georgian language support for modern Information and Communication Technologies (ICT), Georgian computer terminology, as well as specifics of open source software Georgianization and computer aided translation (CAT) will be discussed in the paper.

Overview and samples of open source internet (Mozilla, Firefox, Thunderbird, etc), office suite (OpenOffice) and other software use in education system and the Globalization (G11n), Internationalization (I18n) and Localization (L10n) related tasks, problems and solutions will be presented.

Also, based on the concrete examples of software Georgianization and corresponding terminology by using of glossary of terminology and computer aided translation possibilities will be demonstrated. In particular, one of the examples will be OpenOffice Georgianization technology via translation software especially created for this purposes.

D. Tvaltvadze, S. Sarjveladze

I. Javakhishvili Tbilisi State University

darejantsu@posta.ge

DIGITALIZATION OF THE OLD GEORGIAN TEXTS (RESULTS OF THE SCIENTIFIC GROUP WORKING ON THE CRITICAL EDITION OF THE GEORGIAN VERSIONS OF THE GOSPELS)

Old Georgian original and translated literary texts are digitalized in Georgia and abroad. Prof. Zurab Sarjveladze created the group of scientists to study the critical text of the Georgian translation of Gospels. The group members: Zurab