

II კონფერენცია

ბუნებრივ ენათა დამუშავება

ქართული ენა და კომპიუტერული

ტექნოლოგიები



Conference II

Natural Language Processing

The Georgian Language and Computer Technologies

თბილისი 2004 Tbilisi

საქართველოს პრეზიდენტთან არსებული
სახელმწიფო ენის მუდმივი სახელმწიფო კომისია

საქართველოს მეცნიერებათა აკადემიის არნ. ჩიქობავას
სახელობის ენათმეცნიერების ინსტიტუტი

ბუნებრივ ენათა დამუშავება ქართული ენა და კომპიუტერული ტექნოლოგიები

2004 წლის 21-23 ივნისი

კონფერენციის მასალები



„ქართული ენა“
თბილისი
2004

შუშაობის გეგმა

ორშაბათი, 21 ივნისი

11.00 გ. კვარაცხელია — შესავალი სიტყვა

11.30 თ. უთურგაიძე, ლ. ეზუგბაია, თ. ვაშაკიძე, მ. მანჯგალაძე,
ქ. კაკიტაძე — ერთი და იგივე აქტანტი ორპირიან აქტივში

12.00 ლ. ჩხაიძე — პარადოქსული ტექსტების როლი მეცნიერული
ცოდნის აღწერაში

12.30 ლ. ჩხაიძე — ქართული პარადოქსული ტექსტის ზოგიერთი
მაგალითი

130-14.00 — შესვენება

14.00 მ. ჩაჩანიძე — ენის ლექსიკოგრაფიული მოდელი და ლინგვის-
ტურ მონაცემთა ბაზები

14.30 ქ. ფხაკაძე, გ. ჭანკვეტაძე, ლ. ტიბუა მ. ივანიშვილი,
ვ. ლეკიაშვილი, ე. სოხელია — ბუნებრივი ენათა დამუშავების თანამედროვე
პრობლემატიკა და ქართული ენის პირდაპირი ფორმალურ-ლოგიკური
აღწერა

15.00 გ. ღვინევაძე — არსზე ორიენტირებული ლინგვისტიკა

15.30 რ. ბერულავა — ქართული ენის სტრუქტურის სიტყვათშემოკლე-
ბითი თვისების გამოყენება კომპიუტერულ ტერმინოლოგიაში

სამშაბათი, 22 ივნისი

11.00 ქ. ანთიძე, დ. მიშელაშვილი — ბუნებრივი ენის მორფოლო-
გიური ანალიზატორი

11.30 ქ. ანთიძე, დ. მელიქიშვილი, დ. მიშელაშვილი — ქართული
ენის კომპიუტერული მორფოლოგია

12.00 ლ. ლორთქიფანიძე — ფორმა \leftrightarrow კომპოზიტი \leftrightarrow კონსტრუქცია \leftrightarrow
იდიომი (ავტომატური თარგმნის საკითხისათვის)

12.30 ლ. მარგველანი, ნ. ჭავაშვილი — ქართული ენის მორფო-
ლოგიური ცნობარის კომპიუტერული ვარიანტი

13.00-14.00 — შესვენება

14.00 ბ. სოლოლაშვილი, ჭ. ინგია — ქართული ტექსტის კომპიუტერული გახმოვანება (პრობლემები და მათი ვადაწყვეტის გზები)

14.30 გ. ჩიჩუა — სამეტყველო ფრაზის სიტყვებად დაყოფა და მეტყველების გამოცნობა

15.00 გ. ჭავჭავაძე, კ. ფხავაძე, ლ. ტიბუა მ. ივანიშვილი, ვ. ლეკაშვილი, ე. სოხელია — ქართული ენის ბირთვული ნაწილის კომპიუტერული მოდელი და მარტივი წინადადების სინტაქსური სინთეზატორი და ანალიზატორი

ოთხშაბათი, 23 ივნისი

11.00 რ. მირიანიშვილი — პუმანიტარულ მეცნიერებათა კომპიუტერული მონაცემების ინტერნეტული ინტეგრირება

11.30 ნ. ხუციშვილი — ქართული საისტორიო საბუთების ლოგოსმეტყველური ნაწილის (ინვოკატიო) კომპიუტერული ანალიზისათვის

12.00 ე. დადუნაშვილი, რ. ხეირტლაძე — კომპიუტერული პროგრამა „DS — Folklore 2.0“ და ქართული ფოლკლორის მონაცემთა კომპიუტერული ბანკი

12.30 ო. კაპანაძე, ი. გაბუნია — ონტოლოგიური სემანტიკა ქართული ენის ტექნოლოგიისათვის

13.00-14.00 — შესვენება

14.00 კ. ორჟოშნელი — კომპიუტერული ტერმინოლოგიის ქართულად თარგმნასთან დაკავშირებული პრობლემები

14.30 ქ. დათუკიშვილი, ნ. ლოლაძე, მ. ზაკალაშვილი — ქართული ენის კორპუსი (მიზნები და მოცუანები)

15.00 ნ. ლოლაძე — მნიშვნელობის ანალიზი და კომპიუტერული ლექსიკოგრაფია

სამეცნიერო ლიტერატურაში პირველი პირის ერთპირიან პასივს კონვერსიულად მიემართება აქტივის ოთხი ფორმა (ა. შანიძე).

თუ აქტივის პარადიგმაში 11 კომბინაციაც შევა, მაშინ პირველი პირის ერთპირიან პასივს მიემართება აქტივის 5 ფორმა, კერძოდ: 11, 21, 31, 2¹, 3¹.

რაც შეეხება კომბინაციებს 33 და 3³, თითოეული შეიძლება ორნაირად წარმოდგეს: (1) 3= და =3 (ასევე 3³= და =3³) ერთი და ოგივე აქტანტია (მაგ., მან თავისი თავი ჩაწერა სიაში) და (2) 3= და =3 (ასევე 3³= და =3³) სხვადასხვა აქტანტებია (მაგ., მან მეზობელი ჩაწერა სიაში). პირველ შემთხვევაში კონვერსიულ პასივს არ დაურთვის უბრალო დამატება: მან თავისი თავი ჩაწერა სიაში ~ ის ჩაწერა სიაში; მეორე შემთხვევაში უბრალო დამატება აზუსტებს ვითარებას ქმედების ჩამდენზე მითითებით: მან მეზობელი ჩაწერა სიაში ~ მეზობელი მის მოერ /მისგან ჩაიწერა სიაში.

**კონსტანტინე ფხაკაძე, ველა ჭანკვეტაძე, ლალი ტიბუა
თსუ ი. ვეკუას სახ. გამოყენებითი მათემატიკის ინსტიტუტი,
ლოგიკისა და ენის ქართული ჯგუფი
მარინე ივანიშვილი, ვლადიმერ ლექიაშვილი, ეთერ სოსელია
საქართველოს მეცნიერებათა აკადემიის გ. წერვილის სახ. აღმოსავა-
ლეთმცოდნეობის ინსტიტუტი, ლოგიკისა და ენის ქართული ჯგუფი
pkhakadz@viam.hepie.du.ge**

**ბუნებრივ ენათა დამუშავების თანამედროვე
პრობლემატიკა და ქართული ენის პირდაპირი
ფორმულარულ-ლოგიკური აღწერა**

ლ. ვიტგენშტეინის მიხედვით ლოგიკა არის ენა. რას ნიშნავს ეს? როგორ უნდა გავიგოთ ტერმინი „ენის ლოგიკა?“ საზოგადოდ, ლოგიკა სწორედ აზროვნების წესებია, აზროვნება კი შეიძლება გაგებულ იქნეს როგორც გარკვეული ენობრივ-აზროვნებითი პროცესები. მათემატიკური ლოგიკის მიზანი იყო და არის ამ პროცესთა მკაცრი მათემატიკური ფორმალურიზაცია. ამ მიმართულებით კვლევა ლოგიკაში ფრეგესეული ფორმალურ-მათემატიკური ენის საფუძველზე ვითარდება.

იმისათვის, რომ ანალოგიური ხასიათის კვლევა ვაწარმოოთ ბუნებრივ ენობრივ სისტემათა ფარგლებში, აუცილებელი იქნება მანამდე ამ ბუნებრივი ენობრივი სისტემების აღქვეთური ფორმალური ენების მოძიება.

ბუნებრივი ენის სიტყვების პირდაპირი ფორმალურ-ლოგიკური აღწერა შესაძლებლობას იძლევა მოვიძიოთ მისი იზომორფული ბუნებრივი ფორმალური ენა და ამ ბუნებრივ ფორმალურ ენაში მოდალურ-ლოგიკური შინაარსის სიტყვების სრული ფორმალურ-სემანტიკური ანალიზის საფუძველზე განვავითაროთ ამომწურავი ლოგიკური ხასია-

თის კვლევები, მათ შორის, ფორმალურ მტკიცებათა თეორიის და ე.წ. ფორმალურ-ინტუიციურ (ანუ ზედაპირულ) მტკიცებათა თეორიის შექმნის მიხნით, რაც საფუძვლად უნდა იქცეს უკვე ლოგიკური პროგრამირებით საზღვრული მიღეომების არსებულ ფორმალიზმში გასავითარებლად.

ბუნებრივი ენის სიტყვები აღიწერება მათი ადგილიანობის, ადგილის ტიპისა და სიტყვის, როგორც ოპერატორის მნიშვნელობათა ტიპის მახასიათებელი ინდექსებით. ამგვარად აღწერილი სიტყვების საფუძვლზე ისაზღვრება ენის სწორი ენობრივი გამოსახულების ცნება. სწორი ენობრივი გამოსახულებების შემდგომი კლასიფიცირებით ჩვენ ვდებულობთ წინადადების, არასრული წინადადების, სახელური ფრაზის, ზმური ფრაზისა და შერეული ფრაზის ფორმალურ ცნებებს. ამასთან, ერთამანეთისგან განვარჩევთ სწორი ენობრივი გამოსახულების ფორმალურ-მათემატიკურ და სტანდარტულ ენობრივ ფორმებს. კავშირი ერთი და იმავე სწორი ენობრივი გამოსახულების მათემატიკურ და სტანდარტულ ენობრივ ფორმებს შორის ხორციელდება მექანიკურად, პირდაპირი ჩასმის ოპერატორის შეშვებით.

ბუნებრივი ფორმალური ენის სიტყვების სიმბოლოლური და ლოგიკურ-სემანტიკური ინტერპრეტაციები შესაძლებლობას იძლევა კიდუვ უფრო და ზუსტდეს ბუნებრივ ენობრივ სისტემებში არსებული ე.წ. წმინდა ლოგიკური სიტყვების: 'და', 'თუ', 'ან' ... მნიშვნელობების ფორმალურ-ოპერატორული ბუნება და განხორციელდეს მათი მკაცრი ფორმალურ-მათემატიკური შესწავლა.

ბუნებრივი ენის ფორმალიზაციის ჩვენეული მიღგომა განსხვავდება როგორც 6. ჩომსკის უნივერსალური ლინგვისტურ-გნეტიკური პროგრამის იდეიისაგან, ისე რ. მონტეგიუს მცდელობისაგან, სწორი ენობრივი გამოსახულებების აბსტრაქტორების გზით მოქმია მათი ზოგადი ინტერპრეტირებადი ფორმალური სახეები. ამისათვის იგი იყენებდა ჩორჩის A-ოპერატორს და ფრეგას კომპოზიციურობის პრინციპის მკაცრი ფორმალურ-მათემატიკური ადეკვატის საფუძველზე ახერხებდა ენობრივი გამოსახულებების სინტაქსურ კონსტრუქციებთან მათი სუმანტიკის მიკავშირებას.

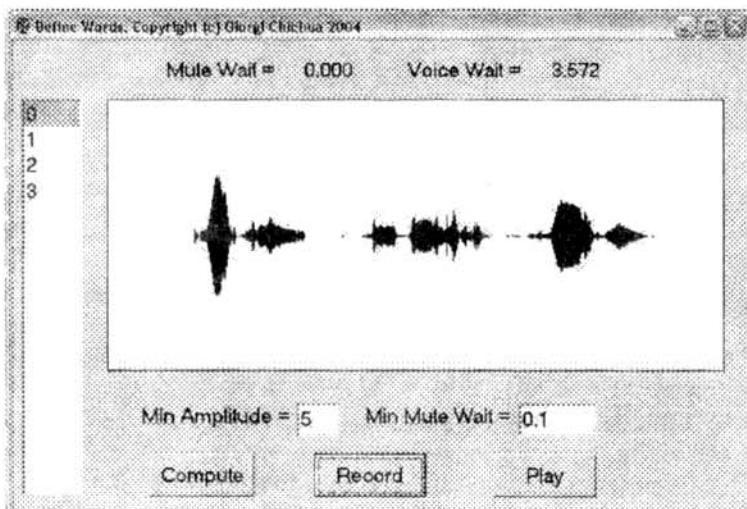
ჩვენი მიღგომებით, ასევე ვდებულობთ სწორი ენობრივი გამოსახულებების განზოგადებულ (აბსტრაქტორებულ) ფორმალურ ინტერპრეტირებად ფორმებს, თუმცა მონტეგიუსგან განსხვავდით, ჩვენთან ეს ხდება არა ენობრივი გამოსახულებების A- აბსტრაქტორებით, არამედ უშუალოდ სიტყვითი მასალის პირდაპირი ფორმალური აღწერის გზით, რის საშუალებასაც ქართული ენის სინტაქსურ-მორფოლოგიური სტრუქტურა იძლევა.

გიორგი ჩიჩუა

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

სამეტულო ფრაზის სიტუვებად დაყოფა და მეტულების გამოცნობა

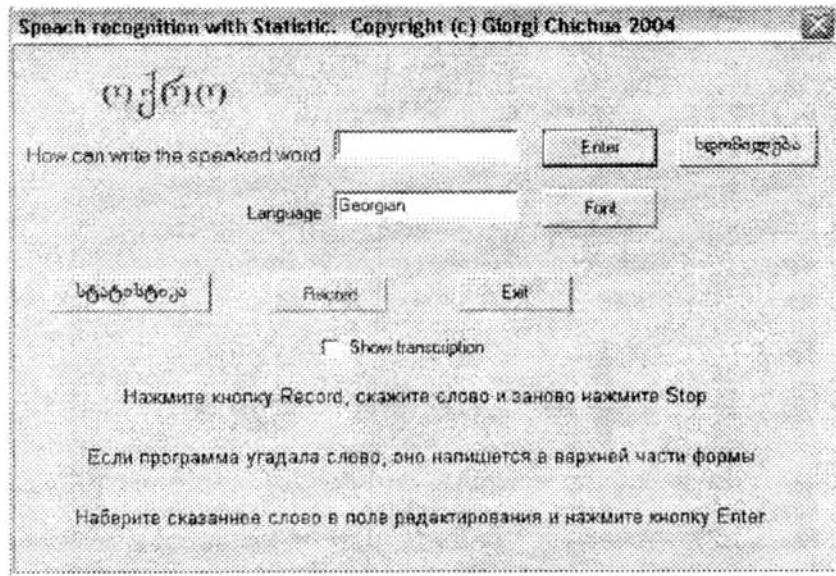
ჩვენ მიერ შემუშავებულია კომპიუტერული პროგრამა, რომელიც
იდეურად და დიზაინურად დაიგეგმა ლოგიკისა და ენის ქართულ
ჯგუფში (ხელმძღვანელი კ. ფხაკაძე). პროგრამა უზრუნველყოფს წინა-
დადების (ფრაზის) შემადგენელ სიტყვებად დაყოფას. პროგრამის ფორ-
მაში უნდა მივუთითოთ **Min Amplitude** მინიმალური ამპლიტუდა, რო-
მელზე ნაკლები სიგნალიც ითვლება სიჩუმედ და ასევე უნდა მივუთი-
თოთ **Min Mute Wait** მნიშვნელობა, რომელიც განსაზღვრავს, მინიმა-
ლურად რა ხანგრძლივობის სიჩუმის ინტერვალი უნდა ჩაითვალოს
სიტყვების გამყოფად. ფორმაზე არის სამი ღილაკი: **Compute, Re-
cord, play.** ხმის ჩასაწერად უნდა დაგაჭიროთ ღილაკს **Record**, უნდა
ვთქვათ წინადადება(ფრაზა) და დაგაჭიროთ **stop**-ს. ამის შემდეგ წინა-
დადება (ფრაზა) ავტომატურად დაიყოფა სიტყვებად და ამ სიტყვების
ნომრები ჩაიწერება პროგრამის ფორმის მარცხნივ არსებულ სიაში. 0
არის მთელი წინადადების ვარიანტი, ხოლო მომდევნო რიცხვები 1 და
ა.შ. არის გამოყოფილი სიტყვები. ცალკეული სიტყვების მოსამა-
ნად უნდა მოვნიშნოთ მაუსით სიტყვის ნომერი და დაგაჭიროთ ღი-
ლაკს **Play**.



პროგრამის ფორმა ფრაზის „კაცი მიდის გზაზე“ წარმოთქმის შემდეგ.

ჩვენ შეგვიძლია ხმის ჩაწერის შემდეგ შევცვალოთ **Min Amplitude** და **Min Mute Wait** მნიშვნელობები, რისი უფასებოც გამოხნდება **Compute** ღილაკზე დაწოლის შემდეგ. პროგრამა ასევე ითვლის ცალკეული გამოყოფილი სიტყვის წინ არსებული სიჩუმის ხანგრძლივობას და სიტყვის ხანგრძლივობას, რასაც წერს ეკრანის ზედა ნაწილში **Mute Wait, Voice Wait.** ეკრანის ცენტრალურ ნაწილში იხაზება სიტყვის შესაბამისი ხმის გრაფიკი.

მეორე წარმოდგენილი პროგრამის დანიშნულებაა წარმოთქმული სიტყვების წერით ფორმაში გარდაქმნა, გამოცნობა. პროგრამა მუშაობს ორ რეჟიმში: სწავლების რეჟიმსა და გამოცნობის რეჟიმში. სიტყვის სწავლებისათვის უნდა დავაჭიროთ ღილაკს **Record**, უნდა კოტვათ სიტყვა და დავაჭიროთ **Stop** -ს. შემდეგ რედაქტირების ველში უნდა ჩავწეროთ წარმოთქმული სიტყვა და დავაჭიროთ **Enter** -ს. გამოცნობის რეჟიმში ასევე უნდა წარმოვთქვათ სიტყვა და მისი გამოცნობის შემთხვევაში ეს სიტყვა დაიწერება ეკრანზე. პროგრამა აღჭურვილია სტატისტიკური დამუშავების მექანიზმით. ღილაკზე “ხდომილება” დაწოლის შემთხვევაში უნდა მივუთიოთ, რა მოხდა: სიტყვის სწავლება, გამოცნობა, შეცდომა თუ ვერგამოცნობა. სტატისტიკის სანახავად უნდა დავაჭიროთ ღილაკს “სტატისტიკა”. პროგრამის ტესტირების შედეგად დადგინდა, რომ 30 სიტყვის სწავლების შემდეგ, გამოცნობის პროცენტი 80% - ზე მეტია.



ნახაზე ჩანს პროგრამის ფორმა, სიტყვა “ოქრო”-ს გამოცნობისას.

ფუძეებით შეცვლას ახორციელებს მთლიანი წინადაღების სემანტიკის დემონსტრაციული იგნორირებით; ტექსტში გამოყენებული გრამატიკული მარკერები უცვლელი რჩება; აღსანიშნავია, რომ შინაარსის მქონე სიტყვების გამოყენებამ არა თუ დაუახლოვა წინადაღება რამე რეალურ სცენას, არამედ მისი ადქმისას ერთგვარი შინაგანი კონფლიქტიც კი გამოიწვია:

მაცხვარებ ჩაზიჯზიპებულში იღლიები შეიძობებს ბააბრიუპებინა

3.10. და ბოლო ნაბიჯიც – 3.9-ში აგებული წინადაღების კორექტული სემანტიკის მიღწევა დექსიკის სათანადო შეცვლით; პრატჩიოკულ მეტყველებაში ამ წინადაღების აგება ხორციელდება კონკრეტული სცენის აღწერის მოთხოვნით, რაც ფორმალური ენის კონსტრუქციების გენერაციის საწინააღმდეგოა; მაგრამ ტექსტის აგების შემდეგ უკვე შეუძლებელია გარჩევა, საქმე გავაქს თუ არა ფორმალურად აგვიტულ კონსტრუქციასთან, რომელსაც სავსებით რეალური ინტერპრეტაცია აქვს, თუ იგი მეტნაკლებად აქტუალური შინაარსის მქონე სცენის აღწერაა ქართული ენის მატარებლის მიერ; ამ გაგებით, აქ სრულდება „მანქანური და ადამიანური ინტელექტების“ ტიურინგისეული კრიტერიუმი:

მაცხვარებ ჩამოძიებულში ღორები ძიზიებულებს ბაატჩავებინა

გელა ჭანკვეტაძე, კონსტანტინე ფხაკაძე, ლალი ტიბუა

ი. ვაკუას სახ. გამოყენებითი მათემატიკის ინსტიტუტი,
ლოგიკისა და ენის ქართული ჯგუფი

მარინე ივანიშვილი, ვლადიმერ ლეგიაშვილი, ეთერ სოსელია

საქართველოს მეცნიერებათა აკადემიის გ. წერეთლის სახ. აღმოსავალეთმცოდნეობის ინსტიტუტი, ლოგიკისა და ენის ქართული ჯგუფი
pkhakadz@viam.hepie.du.ge

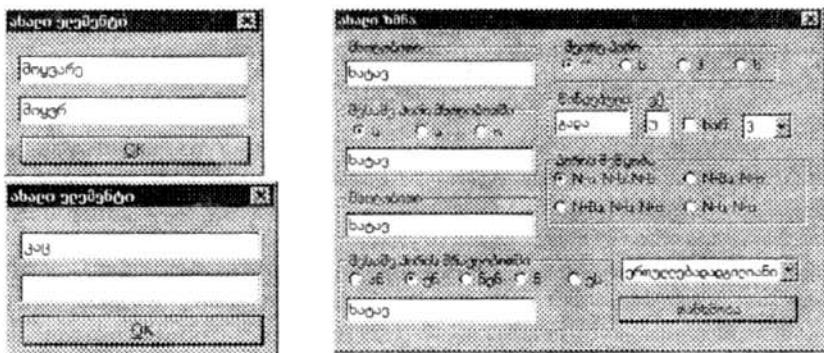
**ქართული ენის ბირთვული ნაწილის კომპიუტერული
მოდელი და მარტივი წინადაღების სინტაქსური
სინთეზატორი და ანალიზატორი**

სისტემა აგებულია ტიპიზებული დექსიკური ბაზის საფუძველზე შემდეგნაირად: არსებითი სახელები, ზედსართავი სახელები, რიცხვითი სახელები, კუთვნილებითი ნაცვალსახელები, ადამიანთა სახელები, წოდებები და ზმები.

სახელების ბრუნების კომპიუტერული მოდელირებისას საპრუნებელი ფუძის შესატანად გამოყოფილია ორი ველი, რისი მეშვეობითაც მარტივად და ბუნებრივად იქნა გათვალისწინებული ქართულ ენაში სახელთა ბრუნების ძირითადი თავისებურებანი. კერძოდ: ა) ხმოვანფუძიან და თანხმოვანფუძიან სახელთა სხვადასხვაგვარი ბრუნება; ბ) კუმშვად სახელთა ბრუნების სპეციფიკა; გ) კვეცად სახელთა ბრუნების სპეციფიკა; დ) კუმშვად-კვეცად სახელთა ბრუნების სპეციფიკა; ე) ფუძის თავისებური ცვილებები. სისტემა ითვალისწინებს აგრეთვე

სინტაქსურად შეკავშირებულ სახელთა ბრუნების თავისებურებებს. მაგალითად, დამოუკიდებლად ნაბრუნები ზედსართავი და რიცხვითი სახელებისა და კუთვნილებითი ნაცვალსახელების ფორმები განსხვავდებულია მათი არსებით სახელთან სინტაქსურ-ოპერატორულად შეკავშირებული ბრუნვის ფორმებისაგან, რაც სისტემაში ტიპიზებულია შესაბამისი ფორმალური კონსტრუქციებით. სისტემა ითვალისწინებს აგრეთვე დანართიან სახელთა ბრუნვის თავისებურებებს. რიცხვით სახელთან სინტაქსურად შეკავშირებული არსებითი სახელის ბრუნებისას ავტომატურად გამოირიცხება მისი მრავლობითი რიცხვის ფორმები.

რაც შეეხება ზმნურ ფორმებს, მათი წარმოებისას პროგრამა კერდნობა კ. ფხაკაბის მიერ შემუშავებულ მეორედებს. კერძოდ პროგრამაში ზმნური ფორმების საკლასიფიკაციოდ და საღერივაციოდ გამოყენებულია კ-, მ- და ვ-ვარ ტიპის ერთადგილიანი და ორადგილიანი უდლებითი სქემები. ამასთან, პროგრამულ-ექსერიმენტული ინტერესებიდან გამომდინარე, კვლევის მიმდინარე ეტაპზე მოხდა ამ უდლებითი სქემების ერთგვარი განზოგადება-გაერთიანება, რამაც თითოეული წინარე ზმნური სემანტიკური ერთეულის საფუძვლზე დერივაციულად წარმოვბადი ყოველი კონკრეტული სრული უდლებითი მწკრივის მისაღებად მოგვთხოვა ამ წინარე ზმნური სემანტიკური ერთეულის ერთიდან მაქსიმუმ ოთხამდე ალომორფული ლინგვისტური რეპრეზენტაცია. ამასთან, თითოეული წინარე ზმნური სემანტიკური ერთეული სისტემაში შემოდის ძირეული ლექსიკური ერთეულის სახით, როგორც ფრეგესეული ადგილიანობითად დახასიათებული და პრედიკატულად განსაზღვრული მონაცემი, რომლის საპრედიკატო ადგილები ტიპიზებულია -ი, -მა, -ს, -ის, -ში და -ზე ტიპის აბსტრაქტული ფორმალური ლინგვისტური კონსტრუქციებით. ამგვარი მიღეომებით შედგენილი ლექსიკონი იძლევა იმის საშუალებას, რომ სისტემამ პროგრამულად უზრუნველყოს უდლებითი მწერივის ყველა დერივაციული ზმნური



ფორმა, და ამასთანავე, თითოეული პროგრამულად ნაწარმოები დერივაციული ზმნური ფორმა გადაიაზროს ფრეგესეული ადგილიანობით

დახასიათებულ და ლინგვისტურად ტიპიზებულ პრედიკატულ სიმბოლო-ნიშნად, რაც საბოლოო ჯამში, მანამდე უკვე -ი, -მა, -ს, -ის, -ში და -ზე ფორმალური ლინგვისტური კონსტრუქციებით ტიპიზებული სახელური ფრაზების საფუძველზე სისტემას აძლევს სინტაქსურად გამართული მარტივი წინადადებების სინთეზირების შესაძლებლობას, და ამავდროულად, სისტემა საშუალებას იძლევა მარტივ წინადადებებად მონიშნული გამოსახულებების სინტაქსური ანალიზისა, რის საფუძველზეც სისტემაში პროგრამულად რეალიზდება ანალიზების შედეგად არაკორექტულად მონიშნული წინადადებების სინტაქსური სპელჩეკირება.

დასკვნის სახით: აქ ნაწილობრივ აღწერილი პროგრამა პირველი, ჯერ მხოლოდ სამოვანი ექსპერიმენტია „ლოგიკისა და ენის ქართული ჯგუფის“ ფარგლებში განვითარებული ფორმალური, ლოგიკურ-ლინგვისტური ხასიათის კალევების კომპიუტერული რეალიზაციისა არაიმპერატიულ პროგრამულ გარემოში. მიუხედავად იმისა, რომ პროგრამის ძირითადი მიზანი არ არის ქართული ენობრივი მასალის სინტაქსური სპელჩეკირება, ჩვენ მივიჩნევთ, რომ წინამდებარე პროგრამა, როგორც ქართული ენობრივი მასალის კომპიუტერული სპელჩეკირების ამოცანის ერთ-ერთი პირველი პროგრამული რეალიზაცია, მნიშვნელოვანია იმითაც, რომ პროგრამის, როგორც ზოგად-თეორიული, ლინგვისტურ-ლოგიკური საფუძველი, ისე მისი ეს კონკრეტული პროგრამული რეალიზაცია ისეა აგებული, რომ თავიდანვე გათვალისწინებული უკვე მოხაზული შედეგების თანახმანით გასრულებისა და დაზუსტების თავისუფალი შესაძლებლობა.

ნიკო ხუციშვილი

საქართველოს მეცნიერებათა აკადემიის კ. კეკელიძის სახ.

ხელნაწერთა ინსტიტუტი

manuscript@iatp.org.ge

ქართული საისტორიო საბუთების ღვთისმეტყველური ნაწილის (ინვოკატიო) კომპიუტერული ანალიზისათვის

ქართული საისტორიო საბუთების შესავალი ნაკვეთის (სიგელის თავის) ერთ-ერთ ძირითად ნაწილს წარმოადგენს ე. წ. ღვთისმეტყველური შესავალი – ღვთისა და წმინდანთა მოხსენიება – „invocatio“.

აღრეულ საბუთებში (XI-XIV სს.) „ინვოკატიო“ შედარებით მარტივია. მოგვიანებით კი, XV ს-დან, ღვთისმეტყველური შესავალი იზრდება და ზოგიერთ შემთხვევაში საქმაოდ კრცელია. მკვლევართა მიერ გამოთქმულია მოსაზრება, რომ გარკვეული სახის საბუთებისათვის (მაგ., ეკლესია-მონასტრებისთვის გაცემულთათვის) „ინვოკატიო“ შედარებით ფართო იყო, ხოლო ზოგიერთი საბუთისათვის (მაგ., კერძო პირ-

few nor discrepant. Despite the abundance of opinions explanatory dictionary of a common language represents the central type.

This conclusion is rather significant. First of all, it confirms the fact that the ideology of linguistic modeling, development tendencies of modern theoretical linguistics and evolution of lexicography proper had been evolving in one direction. That is best of all shown by concepts of lexicographic relevance and linguistic portrayal as well as types of lexicographic information and lexicographic parametrization of a language (in a word – article peculiarly condensed and focused). We reckon that an explanatory dictionary of a common language contains (or represents) a lexicographic model of that language. That is why first of all majority of dictionaries are compiled on the basis of an explanatory dictionary containing dozens of lexicographic parameters. For example, a lot of dictionaries have been compiled and many more will possibly be compiled on the basis of Georgian explanatory dictionary. We have the similar situation in case of explanatory dictionaries of other languages. Seemingly, here we come across a kind of linguistic (resp. lexicographic) universality which is not affected by the fact that the amount of lexicographical parameters in various languages differs. Accordingly, forms of linguistic database (in a conditional and simple way: lexical and grammatical data sources) design and presentation (e.g. network and hierarchical forms as well as electronic dictionaries compiled considering relational scheme) indicate organic and logical links that exist between lexicographic models and linguistic databases.

G. Chankvetadze, K. Pkhakadze, L. Tibua

I.Vekua Institute of Applied Mathematics

pkhakadz@viam.hepi.edu.ge

M. Ivanishvili, L. Lekashvili, E. Soselia

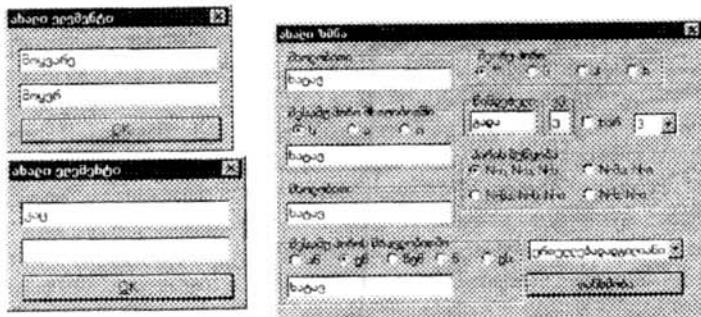
G. Tsereteli Institute of Oriental Studies, Georgian Group of Logic and Language

A COMPUTERIZED MODEL OF THE BASIC PART OF THE GEORGIAN LANGUAGE; SYNTACTIC SYNTHESIZER AND ANALYZER OF A SIMPLE SENTENCE

On the basis of a typified lexical database, the system is designed in the following way: nouns, adjectives, numerals, possessive pronouns, personal names, vocatives and verbs.

In a computerized modeling for name declension two fields are provided for entering the stems that have to be declined, thus basic specifics of name declension in Georgian have been easily considered, these specifics include: a) different kinds of declension for stems ending in vowels and the ones ending in consonants; b) specifics of the contracted noun declension, c) specifics of the clipped noun declension, d) other changes in stems. The given system includes specifics of the declension of nouns that are syntactically joined. For instance, forms of independently declined adjectives and those of numerals and possessive pronouns

differ from declension forms of their syntactic-operational joining with nouns, and that is typified in the system by according formal constructions. The system as well implies specifics of suffixed name declension. When declining a noun that syntactically comes in pair with a numeral, its plural forms are automatically excluded.



With verbal forms, program brings into use methods worked out by K.Pkhakadze i.e. for the classification and derivation of verbal forms, ქ-, ბ- and ვ- type one space and double space declension schemes are used. Also considering program and experimental factors it became possible at the current stage of the research to merge and generalize those declension schemes. In its turn, in order to receive each concrete fully declined screeve that is derivationally formed on the basis of each pre semantic verbal unit, the mentioned merging required linguistic representation of maximum four instead of one allomorphs from pre-semantic verbal unit. Also each Semantic pre-verbal unit enters the system as a basic lexical unit and as location and predicative a datum defined by Frege, predicative places of which are typified by -ი, -ძ, -ს, -ის, -შ and -ბი type abstracted formal-linguistic constructions. A dictionary designed by considering the given approaches enables the program to electronically provide each derivational verbal form of a declensional screeve as well as it makes it possible for each derivational verbal form created by this program, being re-thought into a predicative symbol linguistically typified and locationally described according to Frege. All that eventually enables the system on the basis of previously typified by means of formal linguistic constructions of -ი, -ძ, -ს, -ის, -შ and -ბი noun phrases, synthesize syntactically correct simple sentences. At the same time, the system makes it possible to syntactically analyze images marked as simple sentences and on the basis of that syntactic spell check of sentences marked as incorrect as a result of analysis, is programmatically realized in the system.

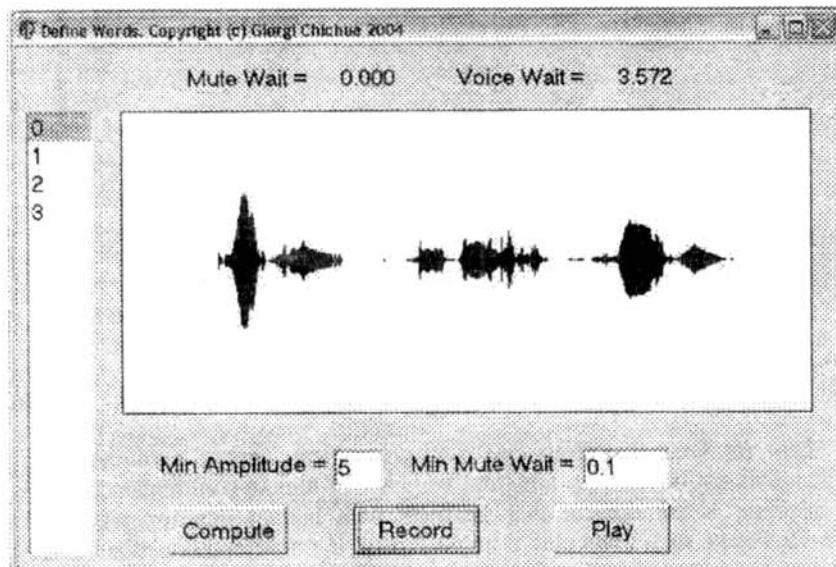
In conclusion: the program partly described here represents just a three months experiment conducted by the Georgian Group of Logic and Language. G.Chankvetadze, a member of the group and researcher is working on the programmatic realization in the prolog system. Although the aim of the program is not syntactic spell check of Georgian linguistic material it is important that the given program right in the beginning aims at further development and perfection.

G. Chichua

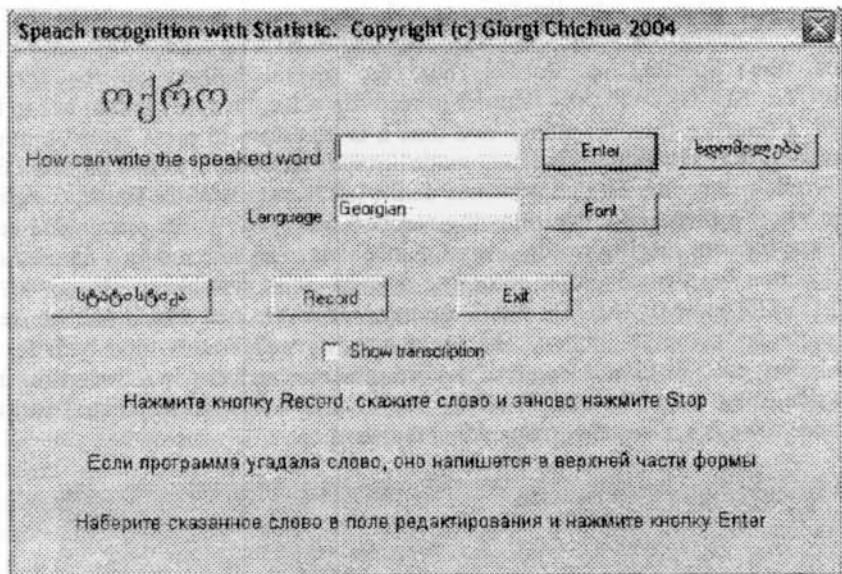
I. Vekua Institute of Applied Mathematics, Georgian Group of Logic and Language
Sulkhan Saba Orbeliani Tbilisi Pedagogic Institute
gchichua@yahoo.com

DIVISION OF A SPEECH PHRASE INTO WORDS AND IDENTIFICATION OF THE DISCOURSE.

We have created a computer programme, design and idea of which has been worked out in the Georgian Group of Logic and Language (Supervisor K. Pkhakadze). The programme insures split of a phrase into its constituent words. We have to indicate in the programme frame MIN AMPLITUDE – minimal amplitude, lesser sound of which is considered silence and also we have to indicate MIN MUTE WAIT as meaning, that would define the minimal interval of silence that has to be applied when splitting up words. There are several buttons on the frame: COMPUTE, RECORD, PLAY. In order to record a voice, "record" button has to be pressed, a sentence (phrase) uttered and then it is necessary to press "stop" button. After that, a sentence (phrase) will automatically split up into words and numbers for those words will be in a list that is located on the left hand side of the frame. 0 stands for the whole sentence and the following numbers are 1 etc. Words are singled out. In order to listen to specific words, a word number has to be selected by a mouse and "play" button be pressed. After recording a voice, it is possible to change MIN AMPLITUDE and MIN MUTE WAIT meanings, the effect of which will be seen after pressing "compute" button. The programme as well counts time both for a pause before each word and length of a word utterance and that is reflected on the upper side of the screen in MUTE WAIT and VOICE WAIT. In the central part of the screen there is a diagram corresponding the word.



The aim of the second presented program is to transform uttered words into written ones, to guess them. There are two modes of operation in the program: teaching mode and identification mode. In order to teach a word, "record" button has to be pressed, a word be uttered and then "stop" button pressed. Further in an editing field, an uttered word has to be typed and "enter" button pressed. In an identification mode, again a word has to be uttered and in case a word is identified it will be written on the screen. The programme is equipped by a statistical processor. By pressing "activity" we have to indicate which process had been undertaken: word teaching, word identification, error or failure to identify. In order to view statistics, "statistics" button has to be pressed. As a result of programme testing the data received showed that after teaching 30 words, identification percentage exceeds 80%.



The drawing shows programme frame when identifying the word "oqro."

L. Chkhaidze

Georgian Academy of Sciences, Arn. Chikobava Institute of Linguistics
levan@dr.com

THE ROLE OF PARADOXICAL TEXTS IN DESCRIPTION OF SCIENTIFIC KNOWLEDGE

1. There are texts that from one sight are logical, correct, written in the language that does not differ from a natural language of human communication and yet their meaning is paradoxical that conflicts with human intuition based on experience. However, such texts play a major role in the development of scientific

K. Pkhakadze, G. Chankvetadze, L. Tibua

I. Vekua Institute of Applied Mathematics

pkhakadz@viam.hepi.edu.ge

M. Ivanishvili, L. Lekashvili, E. Soselia

G. Tsereteli Institute of Oriental Studies, Georgian Group of Logic and Language

MODERN PROBLEMS IN NATURAL LANGUAGE PROCESSING AND DIRECT FORMULAR-LOGICAL DESCRIPTION OF THE GEORGIAN LANGUAGE

According to Wittgenstein language is the logic. What does that mean? How can we understand the term "logic of a language?" Generally, logic represents rules of thinking and thinking could be interpreted as a certain linguistic and thinking process. The aim of mathematical logic has always been to mathematically formalize those processes. The research in logic in this direction is conducted on the basis of Frege's formal-mathematical language.

In order to conduct similar research within systems of natural languages, it is necessary to firstly find formal languages adequate to those natural language systems.

Direct formal-logical description of natural language words enables us to find its natural isomorphic formal language and on the basis of a full formal-semantic analysis of the words of modal-logical content in the given natural formal language, we can develop an exhaustive research of a logical character; also with the aim of creation of formal statement theory and a so called formal-intuitive statement theory, that is to become a basis for development of approaches defined by logical programming into the existing formalism.

Words of natural languages are described according to their location, type of location and word indices. Thus, on the basis of the described words a concept of a correct linguistic portrayal of a language is defined. By a further classification of correct linguistic images we receive formal concepts of sentences, unfull sentences, verbless phrases, verbal phrases and mixed phrases. Also, we distinguish formal-mathematical and standard linguistic forms of a correct linguistic image. The link between similar formal-mathematical and standard linguistic forms of a correct linguistic image is carried out in a mechanical way by means of a direct insertion operator.

Symbolic and logical-semantic interpretations of words of a natural language enable to define so called purely logical words existing in natural language systems, such as "and", "if", "or" etc., as well as define formal-operational nature of meanings and to carry out their accurate formal-mathematical study.

Our approach to the formalization of a natural language differs both from N. Chomsky's idea of universal linguistic-genetic program and R. Montague's effort to find generally interpretable formal images for their correct linguistic images by means of their abstraction. To achieve this, he used Church's λ -operator and on the basis of formal-mathematical adequates of Frege's compositional principle he managed to join their semantics to the syntactic structures of linguistic images.

With our approach we also receive interpretable abstract formal forms of correct linguistic images, although unlike Montague we achieve this not by λ abstraction of linguistic images but by means of a direct formal description of the word material that is possible according to the syntactic-morphological structure of the Georgian language.

B. Sologashvili

Association "Sightless Without Boundaries"

mze@geomail.ge

Z. Injia

"Dachi and Company"

www.dachi.com.ge;

SOUND ATTACHMENT TO GEORGIAN ELECTRONIC TEXTS (PROBLEMS AND SOLUTIONS)

The main aim of sound attachment to electronic texts is to enable sightless people get information independently by means of sounds. Currently, the sightless use books written with Braille. Unfortunately, books prepared in the given way have the following flaws:

- Preparation of Braille books is very costly as well as time consuming.
- Braille books are prepared in one copy and accordingly that hampers their wide dissemination.
- As Braille books are read by finger touches, they get soon worn out and accordingly, impossible to read.
- Those who were not born sightless, have problems with learning reading Braille and accordingly, have no access to such books.

Modern computer techniques make it possible for the sightless to have the old way of getting information replaced by electronic texts that have sound. This can easily solve problems of independent reading and writing for the sightless and with such electronic texts people with visual disabilities can learn the news from web-pages and read both special literature and fiction.

Insonation of a word is carried out by joining separate sounds (phonemes). It is also possible to model a human voice (sounds) on an electronic text, but as the practice has shown, usage of previously recorded and separated segments is much closer to a natural sound. When forming words from segments, it is necessary to consider peculiarities of speech (short pauses, interference, dynamic indicators etc).

Apart from a correct word design, it is important to create a model that would have an intonation typical of natural speech. Correctness of intonation depends on a correct distribution of frequency and amplitude indicators in a sentence. A program of ideal distribution is to carry out the contextual analysis of a sentence.

The programme we have worked on topically distinguishes textual databases and enables a consumer listen to a desired text as well as manage the reading process