

არნ. ჩიქობავას ენათმეცნიერების ინსტიტუტი

ს. შაბაძის, 29 მაისი

საერთაშორისო კონფერენცია „მთა ანტიკონივრეცი და ენობრივი ცვლილებები“  
11.00 (11.00) სასაუბროდ და დასრულებას

დ. ენობრივი — შესავალი სიტყვა  
ა. უაღრესად, დ. ენობრივი — სამართლიან შინაგან თბილისში  
ერთობრივად დასრულებას

ბ. სპირიტუალური, ა. უაღრესად, დ. ენობრივი, მ. მანკადავი,  
ა. უაღრესად, ა. ტრადიციული — ქართული ენის მართლმადიდებლობის  
კონფერენციის დასრულებას

## ბუნებრივ ენათა დამუშავება

გ. კონფერენცია, დ. შაბაძის — ქართული ენის მართლმადიდებლობის  
კონფერენციის დასრულებას

დ. შაბაძის — ქართული ენის მართლმადიდებლობის  
კონფერენციის დასრულებას

2007 წლის 29-30 მაისი

### კონფერენციის მასალები

13.00-14.00 — შესვენება

14.00 — 17.00

ა. უაღრესად, ბ. შაბაძის — ქართული ენის მართლმადიდებლობის  
კონფერენციის დასრულებას

ბ. ლაღაძის — ქართული ენის მართლმადიდებლობის  
კონფერენციის დასრულებას

გ. შაბაძის — ქართული ენის მართლმადიდებლობის  
კონფერენციის დასრულებას

დ. შაბაძის, ე. შაბაძის — ქართული ენის მართლმადიდებლობის  
კონფერენციის დასრულებას

ე. უაღრესად — ქართული ენის მართლმადიდებლობის  
კონფერენციის დასრულებას

დისკუსია

თბილისი  
2007

ლურ კომუნიკაციაში) დადასტურებული ენობრივი მასალა – დაწერილი ტექსტი, წარმოთქმული ფრაზა და სხვ.

ენის აღწერის „კორპუსზე დაფუძნებული“ მეთოდი შეიძლება ჩაითვალოს სტრუქტურული ლინგვისტიკის დესკრიპციული მეთოდის გაგრძელებად და შემდგომ განვითარებად.

ენის აღწერის „კორპუსზე დაფუძნებული“ მეთოდი არის მასალის დამუშავების სტატისტიკური მეთოდი. ამ მეთოდის გამოყენებით მიღებული შედეგების სანდოობა დიდწილად არის დამოკიდებული მონაცემთა რაოდენობრივ მახასიათებლებზე.

ენის კვლევის „კორპუსზე დაფუძნებული“ მეთოდი გამოიყენება ენის ყველა დონის შესასწავლად.

**რატი სხირტლაძე, თედო უთურგაიძე, ლალი ეზუგბაია, მაია მანჯგალაძე, თამარ ვაშაკიძე, თეა ტეტელთაშვილი**  
არნ. ჩიქობავას ენათმეცნიერების ინსტიტუტი  
ენის მოდელირების ასოციაცია  
rati2002@msn.com

## **ქართული ენის მართლწერის შემმოწმებელი პროგრამის საპილოტო ვერსია**

მართლწერის შემმოწმებელი პროგრამის საპილოტო ვერსია ფუძნდება ქართული ენის ფორმალისებულ მოდელზე. მის ბირთვს წარმოადგენს სიტყვაფორმათა წარმოქმნის მოდელი, რომლის ძირითადი კომპონენტებია:

1. ფუძეთა ბაზა;
2. ყალიბების ბაზა;
3. სიტყვაფორმათა წარმოების წესები.

ამჟერად ფუძეთა ბაზაში გვაქვს 97,187 ერთეული. მათგან 6,087 – ზმნური ფუძეა.

არსებული ყალიბები წარმოადგენს ბრუნებისა და უღლების პარადიგმებს. გათვალისწინებულია აგრეთვე ფორმაუცვლელ სიტყვათა ნაწილაკიანი თუ თანდებულებიანი ვარიანტები. გამოყოფილია 330 ზმნური და 70 სახელური ყალიბი.

ფორმაწარმოების წესების საშუალებით ვიღებთ ერთმანეთისგან განსხვავებულ 114,021,470 სიტყვაფორმას. აქედან 2,358,058 ზმნაა. ზმნის ფორმების გენერაციისას ამ ეტაპზე ზმნისწინიანი ვარიანტები არ არის გათვალისწინებული. ზმნისწინთა დართვით ზმნურ ფორმათა რაოდენობა არსებითად გაიზრდება.

ფუძეთა ბაზას საფუძვლად დაედო ქართული ენის განმარტებითი ლექსიკონი (რეატომეული). ბაზის გამდიდრება მოხდა ტექსტური კორპუსის დამუშავების გზით. კორპუსი მოიცავს 436,333 ლექსიკურ ერთეულს, რომელთაგან 294,436 განსხვავებულია. ბაზაში აგრეთვე შეტანი-

ლია ადამიანთა საკუთარი და გეოგრაფიული სახელები, ბოტანიკური ლექსიკა; ასევე გათვალისწინებულია დარგობრივი ლექსიკა.

ტექსტში შეცდომის აღმოჩენის შემთხვევაში მართლწერის შემმოწმებელი პროგრამა მომხმარებელს სთავაზობს სიტყვის საგარაუდო სწორ ფორმებს. პროგრამაში გათვალისწინებულია დასაშვებ უზუსტობათა შესაძლო ვარიანტები. თუ მომხმარებელმა მართებულ ფორმაში არასწორად აკრიფა სამი ან სამზე ნაკლები ასო, დააკლო, დაამატა ან შეცდომით სხვა გრაფემით ჩაანაცვლა გრაფემა / გრაფემები, მას პროგრამა 114,021,470 სიტყვიდან აღნიშნული კრიტერიუმების გათვალისწინებით შესთავაზებს სავარაუდო ვარიანტებს. ბუნებრივია, საძიებო სიტყვაფორმების სიმრავლე მოქმედებს პროგრამის სწრაფქმედებაზე. მომხმარებელს შეუძლია დააჩქაროს პროგრამის მუშაობა, თუ იგი დააფიქსირებს, რომ დასაშვებ შეცდომათა რაოდენობა სამზე ნაკლებია (ორი ან ერთი) ან აჩვენოს, სიტყვაში რამდენი ინიციალური გრაფემა უნდა ჩათვალოს პროგრამამ კორექტულად.

თუ მართლწერის შემმოწმებელი პროგრამა ვერ ამოიცნობს ლექსემას, რომელსაც მომხმარებელი კორექტულად მიიჩნევს, შესაძლებელია ამ ენობრივი ერთეულის პირად ლექსიკონში ჩამატება, რითაც აღმოიფხვრება არსებული ხარვეზი და დაფიქსირებულ სიტყვას პროგრამა მომავალში შეცდომად აღარ ადიქვამს.

მართლწერის შემმოწმებელი პროგრამა ქართული ტექსტებისათვის გავრცელებული ყველა ტიპის კოდირებისთვის მუშაობს. სპეციფიკური კოდირების შემთხვევაში კი მომხმარებელს მარტივად შეუძლია მისი აღწერა, რის შემდეგაც პროგრამა მას ამოიცნობს.

მართლწერის შემმოწმებელი პროგრამის პაკეტს თან ერთვის ისეთი სასარგებლო ინსტრუმენტები, როგორცაა სხვადასხვა კოდირებაში აკრეფილი ტექსტის ან მისი ფრაგმენტის ერთ შრიფტსა და კოდირებაში გადაყვანისათვის აუცილებელი საშუალებები, ასევე სიტყვის სწორად გადატანის წესები.

**თედო უთურგაიძე, ლალი ეზუგბაია**  
არნ. ჩიქობავას ენათმეცნიერების ინსტიტუტი  
lezugbaia@yahoo.com

## **სამპირიან ზმნათა ორი ჯგუფი ორთოგრაფიულ ლექსიკონში**

1. სამპირიან ზმნათა დიდ უმრავლესობაში პირცვალებადია მხოლოდ ირიბი ობიექტი; სამპირიან ზმნათა უფრო მცირე რაოდენობაში კი პირის მიხედვით შეიძლება პირდაპირი ობიექტიც იცვალოს.

2. ამ ორი ჯგუფის სამპირიანი ზმნების გრამატიკულ ფორმათა რაოდენობა განსხვავებულია, რაც ელექტრონულ ორთოგრაფიულ სისტემაშიც უნდა აისახოს. “გრამატიკული ფორმა” ვარაუდობს სიტყვის

**G. Shabashvili**

Tbilisi Iv. Javakishvili State University

Faculty of the Humanities

## **THE REALIZATION OF THE PRESENT PERFECT FORMS FROM THE COMPUTATIONAL PRAGMATICS VIEWPOINT**

Speech act interpretation is a classic pragmatic problem. Nowadays in case of computational linguistics, two distinct computational paradigms must be intended in speech act interpretation: a literary meaning, which is based on a logic-based approach and an indirect meaning, which is based on a probabilistic approach. The data of the computational semantics is also relevant.

Present perfect verbs are not used in case of narration in modern Georgian, for the purpose of this semantics in above-mentioned pragmatic environment analytical perfect forms occur along with the present verbs.

The present perfect forms keep the literary meaning in dialogues often during the answering the questions.

When the present perfect forms are independently used, these verbs have the semantic gradation and receive a proposal meaning, for example the category of evidentiality.

In conclusion, we can emphasize in the pragmatic environment the present perfect forms are realized with the literary or approaching meanings. Likewise, it is possible to make some algorithm, which shows this regularity.

**R. Skhirtladze, T. Uturgaidze, L. Ezugbaia, M. Manjgaladze,  
T. Vashakidze, T. Teteloshvili**

Ann. Chikobava Institute of Linguistics

## **PILOT VERSION OF THE GEORGIAN ORTHOGRAPHY CHECKING PROGRAM**

The pilot version of the Georgian orthography checking program is based on formalised models of the Georgian language. Its basic component is the module of the word-forms producing process which in its turn consists of:

1. Stems base;
2. Patterns base;
3. Rules of word-formation.

97, 187 entities are represented in the stems base and out of them 6,087 are verb stems.

Existing patterns are actually paradigmes of noun declination and verb conjugation. Unchangeable word-forms with particles and pre-or post-positions are also considered. Altogether, 400 patterns are marked out: 330 verb-patterns and 70 noun-patterns.

Operating in accordance with the rules of word-formation, we get 114,021,470 word-forms and 2,358,058 out of them are verbs.

On the given stage, when generating verbs-forms, we don't consider forms with pre-verbs. Adding pre-verbs will greatly increase the number of the verb-forms.

The verbs base is founded on the eight volumed Georgian Language Explanatory Dictionary. The base was enhanced by means of text-corpus processing. Text corpus contains 436,333 lexical entities, 294,436 are different. This base includes also human and geographic proper names, lexis of botanics; vocabularies, related to differen spheres of human life are also taken into consideration.

When finding an error, the checking program suggests to an user to choose some correct forms. Some variants of the possible errors are foreseen in the program. If an user chooses a correct form but makes mistakes in spelling - omits some letters or uses extra-letters, or graphemes, the program can suggest him on the mentioned critica, possible versions, aut of 114,021,470 words.

Naturally, great variety of searched forms influences upon speed of the operation. An user can make the process quicker if he can fix those possible mistakes the number of which is less than three (two or one) or show, how many initial graphemes should be considered correct by the program.

If the checking program is not able to recognize a lexeme, which an user presumes correct, it is possible to add this lexicalentity to the private dictionary. In this way the fault of the program will be eradicated and the program, in the future won't consider the added lexeme incorrect.

The orthography checking program works for all the existing types of the Georgian texts coding. In case of a specific coding, the user can simply describe a text and the program will recognize it.

A number of useful instruments is enclosed to the checking program, such as: the means of transferring a text or its part, typed in different coding system in one type of print and code, or the rules of dividing words when carrying from one line to another.

**B. Tsipuria, D. Barbakadze**

Ministry of Education and Sciences of Georgia

## **THE FIRST ANALOGY OF PROJECT GUTENBERG IN GEORGIA**

1. Deer Leap - program of the Ministry of Education and Science of Georgia - is realizing the project „Vakhtang VI“, that is digitalization of the writings by Georgian and those of foreign writers translated into Georgian. The project aims to prepare e-portal, where Georgian literary texts, as well as foreign literary texts translated into Georgian will be published and updated regularly.

2. The works are carried out with utilization of intellectual and technical resources of schools, involved in Deer Leap program, therefore as an intermediate stage the contest of the writer's web-sites is realized, where schools present the web-

