

არნ. ჩიქობავას ენათმეცნიერების ინსტიტუტი

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

2006 წლის 5-6 დეკემბერი

კონფერენციის მასალები

დად თემის ნიშნით), რომელთაც სავარაუდოდ ახასიათებთ უღლების მსგავსი ნიშნები. თითოეული სახელზმნიდან უნდა გამოიყოს ფუძე, განისაზღვროს შესაძლო ზმნისწინთა ნუსხა, ფორმანაკლებობა და წარმოების წესები, რომელთა გარკვეული ნაწილი შეიძლება საერთო ჰქონდეს სხვადასხვა ზმნასთან.

3) სიტყვაფორმათა მიღებული სია უნდა ჩაიწეროს ძიებისათვის (შედარებისთვის) მოხერხებული ფორმით. რამდენადაც სახელური ფორმები არ განიცდიან საწყისი ფორმისგან ძირეულ ცვლილებას, 2.ა) ფორმაწარმოების პრინციპები შეიძლება ჩაიდოს თვით მაძიებელ პროგრამაში და გამოყენებული იქნეს ძიების პროცესშივე. ზმნების შემთხვევაში კი იგივე პრინციპი საკმაოდ არარენტაბელურია და მიღებული ფორმები წინასწარ უნდა დალაგდეს ანბანური თანმიმდევრობით. მაძიებელ პროგრამას აგრეთვე უნდა შეეძლოს რიცხვითი სახელების, ემვატიკური და სხვა ფორმების დინამიურად წარმოება.

აღნიშნული პრინციპები რეალიზებული იქნა პროგრამულ პაკეტში "აი ია" (1997წ), პროგრამაში "მართლწერის შემოწმება ვორდში" (2000წ) და ქართულ ტექსტებთან სამუშაო ხელსაწყობთა კრებულში "საპოვნელა" (2003წ). ლექსიკონის მოწყობის დეტალური აღწერა განთავსებულია ვებგვერდზე www.dachi.com.ge.

მარიამ მანჯგალაძე, თამარ ვაშაკიძე, თეა ტეტელოშვილი, რატი სნირტლაძე

არნ. ჩიქობავას უნათმეცნიერების ინსტიტუტი

ქართული ენის სახელთა ელექტრონული ვერსია

1. მეტყველების ნაწილთა ელექტრონული ვერსია ემყარება ენაში არსებულ შესაბამის მოდელებს და ითვალისწინებს სათანადო ერთეულთა ფორმაცვალების შედეგად რეალიზებულ ყველა შესაძლო მორფონოლოგიურ ვარიანტს. ამის საფუძველზე გამოყოფილია არსებითი სახელის (საზოგადო და საკუთარი, მათ შორის, გეოგრაფიული სახელების), ხედსართავი სახელის, რიცხვითი სახელისა და ნაცვალსახელის მორფონოლოგიური ტიპები შესაბამისი ბაზებითურთ, რომლებიც ემყარება ქართული ენის განმარტებითი ლექსიკონის რვატომეულს. სახელების (იქნება ეს არსებითი, ხედსართავი, რიცხვითი თუ ნაცვალსახელი) თითოეული ტიპი გარკვეული ნიშნით განსხვავდება დანარჩენი ტიპებისგან, მაგალითად, მრავლობითი რიცხვის ქონა-არქონისა თუ კუმშვასა და კვეცასთან დაკავშირებული თავისებურებებით. ზემოთქმულის გათვალისწინებით, მეტყველების ნაწილთა ელექტრონულ ვერსიაში გვაქვს არსებითი სახელებისა და გასუბსტანტივებულ ხედსართავთა 21, ხოლო ნაცვალსახელთა 30 მორფონოლოგიური ტიპი (ნაცვალსახელის ბრუნებასთან დაკავშირებული მრავალი თავისებურების გამო

ისინი დამოუკიდებელ ჯგუფებად გამოყავით). რაც შეეხება რიცხვით სახელებს, მათი ნაწილი არსებით სახელთა შესაბამის ტიპებში გადა-
ნაწილდა, ნაწილი კი დამოუკიდებელ ჯგუფებად გამოიყო. ნებისმიერი
გადახრა საზოგადო წესებიდან, თუნდაც ის მხოლოდ ერთ ლექსიკურ
ერთეულს ეხებოდეს, ფიქსირდება ელექტრონულ ვერსიაში და ხდება
დამოუკიდებელი მორფონოლოგიური ტიპის გამოყოფის საფუძველი.

2. მომხმარებელს სრული შესაძლებლობა ექნება მეტყველების
ნაწილთა ელექტრონულ ვერსიაში შეამოწმოს მისთვის საინტერესო
სახელის (იქნება ეს არსებითი, ზედსართავი, რიცხვითი თუ ნაცვალსა-
ხელი) როგორც მხოლოდ თანდებულიანი და ნაწილაკიანი, ასევე სა-
ხელზე თანდებულისა და ნაწილაკის ერთად დართვის შედეგად მიღე-
ბული ყველა შესაძლო ვარიანტი. ზემოაღნიშნულიდან გამომდინარე,
რომელიმე მორფონოლოგიური ტიპის ერთი ლექსიკური ერთეული
წარმოგვიდგენს 1600-მდე განსხვავებულ ფორმას (მაგ., კაც-ი-ც, კაც-მა-
ც, კაც-ს-ა-ც..., კაც-ი-ვე, კაც-მა-ვე, კაც-ს-ა-ვე..., კაც-ი-ღა, კაც-მა-ღა,
კაც-ს-ღა..., კაც-ი-მეთქი, კაც-ი-თქო, კაც-ი-ო..., კაც-მა-მეთქი, კაც-მა-თქო,
კაც-მა-ო..., კაც-ი-ვით, კაც-ში, კაც-ზე, კაც-თან, კაც-ის-გან, კაც-ის-კენ...,
კაც-ში-ც, კაც-ზე-ც, კაც-თან-ა-ც, კაც-ის-კენ-ა-ც..., კაც-ში-ვე, კაც-ზე-ვე,
კაც-ის-კენ-ვე..., კაც-ში-ღა, კაც-ზე-ღა, კაც-ში-თქო, კაც-ზე-მეთქი, კაც-
ზე-ო და სხვა მრავალი). თანდებულიანი და ნაწილაკიანი ფორმათა კომ-
ბინაციების ენაში არსებობა შემოწმებულია შესაბამის ბაზებში არსე-
ბული ყველა ლექსიკური ერთეულის მიხედვით.

3. გათვალისწინებულია, აგრეთვე, მსაზღვრელ-საზღვრულის
ბრუნებასთან დაკავშირებული თავისებურებანი, კერძოდ, ნებისმიერი
ატრიბუტული თუ სუბსტანტიური მსაზღვრელის ბრუნება არსებით სა-
ხელთან პირდაპირი და ინვერსიული წყობით. აუცილებელი გახდა სა-
ხელის დაფიქსირება ფუძის სახით, ნულოვანი აღომორფით. მაგ., მე-
გობარ თამაზს გადავეცი წიგნი. შემოწმებულია ყველა საზოგადო არ-
სებითი სახელი (აგრეთვე, ზედსართავი, რიცხვითი სახელი და ნა-
ცვალსახელი), რომელიც შესაძლებელია დადასტურდეს ატრიბუტული
მსაზღვრელის ფუნქციით, მაშასადამე, გარკვეულ ბრუნებებში წარმო-
გვიდგეს ფუძის სახით.

4. ამრიგად, ქართული ენის სახელთა (არსებითის, ზედსართავის,
რიცხვითისა და ნაცვალსახელის) უკვე დასრულებული ელექტრონული
ვერსია მომხმარებელს შესთავაზებს აღნიშნული მეტყველების ნაწილე-
ბის ნორმატიულ ფორმებს.

3) Created word-forms list must be transformed as quick-searching database. As noun and adjective forms can little differ from root form, 2.a word-formative processing can be included and used in searching program. In case of verbs this principle is very ineffective and word-forms must be sorted previously. Searcher program also must be able dynamically formative of word-numbers, emphatic and other forms.

These principles of spell-checking has been realized in the package "AI IA" (1997), "Georgian spelling checker for MS Word" (2000); Georgian text tools - "Sapovnela" (3003). Principles of dictionary are described in details at: www.dachi.com.ge

**M. Manjgaladze, T. Vashakidze, T. Teteloshvili,
R. Skhirtladze**
Institute of Linguistics

ELECTRONIC VERSION OF THE GEORGIAN NAMES

1. Electronic versions of parts of speech are based on appropriate models existing in a language and provide grammatical and morphologic conditions for corresponding units. Morphologic types with their corresponding basis (according to GED) of a noun (naturally, common and proper, among them geographic names), an adjective, a numeral and a pronoun are distinguished. Each type (it can be of a noun, an adjective, a numeral or a pronoun) differs from other types by specific characteristics (e. g. peculiarities connected with having or not having a plural or syncope and elision). Taking into consideration above mentioned, there are 21 types of nouns and substantivized adjectives, 30 types of pronouns (owing to many peculiarities connected with a pronoun declension, we distinguished a pronoun as an independent group). As for numerals, they are distributed in the corresponding types of nouns. Any deviation from general rules (even of one lexical unit) is fixed in an electronic version (and is distinguished as an independent type).

2. An user will have a full chance to check any part of speech (he/she is interested in - it can be a noun, an adjective, a numeral or a pronoun) with possible preposition and particle (also, combinations of a preposition and a particle). One lexical unit of any morphologic type presents about 1600 different forms (e. g. *ჰა-ი-ც*, *ჰა-მა-ც*, *ჰა-ს-ა-ც...*, *ჰა-ი-ვე*, *ჰა-მა-ვე*, *ჰა-ს-ა-ვე...*, *ჰა-ი-რა*, *ჰა-მა-რა*, *ჰა-ს-რა...*, *ჰა-ი-მეტი*, *ჰა-ი-ტო*, *ჰა-ი-ო...*, *ჰა-მა-მეტი*, *ჰა-მა-ტო*, *ჰა-მა-ო...*, *ჰა-ი-ვიტ*, *ჰა-ში*, *ჰა-ზე*, *ჰა-თან*, *ჰა-ის-გან*, *ჰა-ის-გენ...*, *ჰა-ში-ც*, *ჰა-ზე-ც*, *ჰა-თან-ა-ც*, *ჰა-ის-გენ-ა-ც...*, *ჰა-ში-ვე*, *ჰა-ზე-ვე*, *ჰა-ის-გენ-ვე...*, *ჰა-ში-ვა*, *ჰა-ზე-ვა*, *ჰა-ში-ტო*, *ჰა-ზე-მეტი*, *ჰა-ზე-ო* and others). Using (from the having-non-having viewpoint) of a preposition and a particle (also combinations of a preposition and a particle) with this or that part of speech (with a noun, an adjective,

a numeral or a pronoun) is checked according to all lexical unit existing in corresponding basis.

3. Peculiarities connected with declension of a determinandum-determinatum are also taken into consideration (how any attributive or substantive determinandum will behave through declension in the case of direct and inversive structure). Here, naturally it is necessary to fix this or that noun in a form of a stem with a zero allomorph, among them common nouns (e. g. megobar tamazs gadavcci çigni). All common nouns (also adjectives, numerals and pronouns), which can be evidenced with the function of attributive determinandum (thus, it can be in a form of a stem in some cases) are checked.

4. Thus, the work on electronic version of Georgian parts of speech (a noun, an adjective, a numeral and a pronoun) is over, the aim of which is to give a user normative forms.

I. Modebadze, T. Tsitsishvili
Shota Rustaveli Institute of Literature

KARTVELOLOGY SCIENCE TREASURY

1. By the end of the 20th century up-tempo of life and technical progress has radically altered the attitude of the individuals towards informational technologies. One of the vivid examples of the above-mentioned is decrease of the demand for the printed materials replacing them by electronic versions.

2. Time has conditioned necessity for the establishment of the electronic libraries and their constant updating. Their number increases day by day in the internet world space and presumably after a certain time their integrated centralized net will be established.

3. A lot of activities are carried out in Georgia in this' direction (The most experienced in the mentioned field is the Parliament National Library (bases: "catalogue", "classical literature", " Levani", etc), however, they deal only with transferring the fiction into the electronic format. Electronic versions of the scientific production accumulated in Kartvelology during the centuries have not been yet prepared in terms of electronic library database.

4. Transfer of the achievements of the Kartvelology science in a new form and their placement on internet in a form of systematic-structural database seems to us as an urgent task in view of maintaining our cultural treasury in accessible modern form for our future generations.

5. Certain activities have been carried out in our institute, which were required for the preparatory stage needed for the fulfillment of the mentioned task (Three different types of scientific works were chosen by us as a sample-pilot material: 1) Monograph; 2) Collection of scientific works; 3) Annual scientific magazine type collection).



გამომცემლობა „უნივერსალი“

თბილისი, 0128. 0. შავაშვილის რაზმ. 1, ☎: 29 09 60, 8(99) 17 22 30

E-mail: universal@internet.ge

Conclusion: the mathematical translation of the Georgian sentences can be made automatically according to certain mathematical properties of morphemes and between word's empty places.