

არნ. ჩიქობავას ენათმეცნიერების ინსტიტუტი

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

2006 წლის 5-6 დეკემბერი

კონფერენციის მასალები

ლალი ეზუგბაია, თელორე უთურგაიძე, რატი სხირტლაძე
არნ. ჩიქობავას ენათმეცნიერების ინსტიტუტი

ქართული ენის ზმნათა ფორმების ელექტრონული ვერსია

1. ქართულ ენაში გამოყოფენ უღლების ოთხ ტიპს (ა. შანიძე). თითოეულ ტიპს აქვს სხვა ტიპათაგან განსხვავებული უღლების საკუთარი მოდელი. ქართული ენის ზმნის ელექტრონულმა ვერსიამ უნდა სრულად წარმოაჩინოს პირიან ფორმათა ბაზები.

სამეცნიერო ლიტერატურაში განსაზღვრულია მწკრივთა რაოდენობა და ასევე ფორმათა რაოდენობა პირისა და რიცხვის კატეგორიების მიხედვით ერთი მწკრივის ფარგლებში.

2. მწკრივთა რაოდენობა ჩვენს სისტემაში ერთით მეტია (12). დამატებულია ბრძანებითი მწკრივი. პირიან ფორმათა განსაზღვრისას ჩვენთვის არსებითია გამოხატულებისა და შინაარსის პლანების კავშირი, რის შედეგადაც ჩვენს ბაზებში ფორმათა რაოდენობა მწკრივის პარადიგმაში საგრძნობლად განსხვავდება სამეცნიერო ლიტერატურაში მიღებული რაოდენობისაგან, მაგალითად: ორპირიან აქტივთა ერთ მწკრივში, ნაცვლად 18-ისა, არის 32 ფორმა, სამპირიანებში კი - 64 ან 128. ეს უკანასკნელი მაშინ, თუ ზმნას პირცვალებადი პირდაპირი ობიექტიც შეეწყობა. აქედან ისიც ჩანს, რომ ჩვენს სისტემაში დაშვებულია სამპირიანი ზმნებიც.

3. ერთი გრამატიკული მოდელის ფარგლებში გამოიყოფა რამდენიმე (ზოგჯერ ათეული, მაგალითად, (ებ) თემის ნიშნიან ზმნებში 36) მორფონოლოგიური ტიპი, რომელთაგან თითოეული მოლიანობაში განსხვავებულია ყველა დანარჩენისგან, რისი დაფიქსირება აუცილებელია ორთოგრაფიულად გამართულ ბაზებში, მაგალითად: (ა) -აე თემისნიშნიან გაუსმოვანებულ ძირთან ზმნათა ნაწილი წყვეტილში ირთავს -ი სუფიქსს I და II პირებში (დავწვი, დასწვი), მაშინ როცა კლავს ტიპის ზმნები ამ პირობებში სმოვანს იჩენენ ფუძეში და -ი სუფიქსიც არ არის აუცილებელი (მოგკალ // მოგკალი, მოჰკალ // მოჰკალი); (ბ) გაუსმოვანებელი ფუძის მქონე -ებ თემისნიშნიანი ზმნები III სერიაში თემის ნიშანს არ გადაიყოლებენ (აგებს ~ აუგია, სცხებს ~ მოუცხია...), გასმოვანებული ფუძის მქონენი კი -ებ თემის ნიშანს I თურმეობითშიც გადაიყოლებენ (აშენებს ~ აუშენებია) და, განსხვავებით ძველი ქართულისა და არქაული კილოებისგან, III სერიის მომდევნო მწკრივებშიც ირთავენ -ებ სუფიქსს (აეშენებინა, აეშენებინოს) და სხვ.

4. აღნიშნული ოთხი ტიპის ფარგლებში უღლების მორფონოლოგიური ტიპების რაოდენობა, თავისებური უღლების მქონე ზმნათა გათვალისწინებით, ორასს აღემატება.

ელექტრონულ ბაზებში ყველა ნიუანსი გასათვალისწინებელია, რათა მომხმარებელს სრულად მიეწოდოს სალიტერატურო ენის ნორმებით დაშვებული ფორმები.

6. The third variant uses other alternative variants of phonetic correspondences in the Colchian language connected with laringal phoneme interchange, insertion of the sonor "n" and increase in the intensity of action expressed by the preverb "do":

ma/chkhvit'/e/k ina/ch'qon/al/'s ghlomp/ep/i do/t'anch'qan/u

7. The fundamental result is the possibility to solve the above set task (a)-(d) mainly by phonetic means with minimal use of semantically differing morphemes. Despite a certain phonetic ambiguity that does not exceed the limits of phonetic variability of the Kolkhian language for which there are no universally recognized literary norms, morpho-syntactic isomorphism (along with a formalized model of the language) practically fully remains. All of this represents another serious argument in favour of a near kindred between the languages of the Kartvelian linguistic family.

L. Ezugbaia, T. Uturgaidze, R. Skhirtladze
Institute of Linguistics

ELECTRONIC VERSION OF GEORGIAN VERB FORMS

1. Four types of conjugation are distinguished in Georgian (A. Shanidze). Each type has its own model differing from the other one. Electronic version of the Georgian verbs must completely manifest bases of personal forms.

Number of screeves and also number of forms according to person and number categories within one screeve are defined in the scientific literature.

2. There is one more screeve in our system (12). An imperative screeve is added to it. During definition of personal forms connection of levels of expression and content is important for us as a result of which a number of forms in our bases in a screeve paradigm differs from a number of form accepted in the scientific literature; e. g. there are 32 forms instead of 18 in one screeve of bi-personal active voice, in tri-personal - 64 or 128. The latter occur when personal-changing direct object is arranged with a verb. According to this it is clear that tri-personal verbs also occur in our system.

3. Within one grammatical model several (sometimes ten, e. g.: there are 36 types in the verbs with **-eb** theme-mark) morphonologic types are distinguished. Each of them differs from other which is **necessary** to be fixed in orthographical correct bases; e. g. (a) **-n** suffix is added to a part of verbs having vowelless-root with **-av** theme-mark in the 1st and 2nd person in aorist (**დავწვი**, **დასწვი**), while in the verbs of **klavs** type a vowel occur in a stem in this condition and **-n** suffix is not necessary (**მოვკალ** // **მოვკალი**, **მოხკალ** // **მოხკალი**); (b) the verbs having vowelless-stem with **-eb** theme-mark do not preserve a theme-marker in the 3rd series (**აგებს** ~ **აგია**, **სცხებს** ~ **მოუცხია**..), but the verbs having vowel-stem preserve **-eb** theme-mark in Present-Perfect (**აწენებს** ~ **აუწენებია**) and differently

from old Georgian and archaic subdialects **-eb** suffix is added to them in the following screeves of the 3rd series (**aešenebina**, **aešenebinos**), etc.

4. A number of phonological types of conjugation within the limits of the mentioned four types, taking into consideration the verbs having peculiar conjugation, is more than two-hundreds.

All nuance in the electronic bases must be taken into consideration in order to give an user the forms permitted through the literary language normes.

Z. Injia

"Dachi and Company"

GEORGIAN SPELLING CHECKER

Principle of Spellchecker program based on taking word between spaces, parsing them from characters (, . ; : ! ? () { } [] ~) and searching in dictionary. The word is considered as error in case if it was not found. Such program is necessary to create dictionary and fast finder function. Solution of this task can be divided in several stages:

1) Creating dictionary by keyboard inputting or by picking word-forms from texts making a sorting list excluding equal forms; second method is quite faster but some additional job is necessary:

a) Only root forms of nouns and adjectives must be saved (children word-forms must be excluded from the list). If this form does not exist, it must be created manually;

b) Only root forms of verbs must be saved. If this form doesnt exist, it must be created manually. Children word-forms can be saved in other lists as examples for future processing. Irregular verbs (Verbs with undefined core form) must be excluded and saved in other list;

c) Unchangeable word-forms must be excluded and saved in other list.

2) For full word-formative processing from the given list of core forms it is necessary:

a) Each item of the list must be marked with self specifications (vowel syncope and elision - ability, plurality, impossible suffixes...). For solution of this task some subprograms can be used. Word groups with like specifications can be defined automatically. Next step is creating the program for word-formation processing using predefined marks

b) Verb word-formation. Automation of this process is very difficult because there are numerous kind of specifications and their combinations in Georgian. This task take more human resources, but it is possible to combine like verbs in the groups by programs (likeness in parts of declension)



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 mathematical properties of morphemes and between word's empty places.