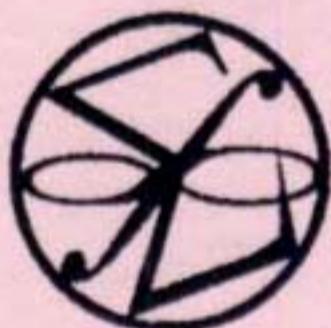


IV კონფერენცია

ბუნებრივ ენათა დამუშავება

ქართული ენა და კომპიუტერული
ტექნოლოგიები



Conference IV

Natural Language Processing

The Georgian Language and Computer Technologies

არნ. ჩიქობავას ენათმეცნიერების ინსტიტუტი

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

2006 წლის 5-6 დეკემბერი

კონფერენციის მასალები

მუშაობის გეგმა

სამშაბათი, 5 დეკემბერი

11.00 **ლ. ეზუგბაია** — შესავალი სიტყვა

11.30 **მ. მანქვალაძე, თ. ვაშაკიძე, თ. ტეტელიშვილი, რ. სხირტლაძე** — ქართული ენის სახელთა ელექტრონული ვერსია

12.00. **მ. ჩაჩანიძე** — ზმნისართის პრობლემა ქართული ენის სრულიად ფორმალიზებული მოდელის აგებისათვის

12.30 **ჭ. ინჯია** — ქართულ ტექსტში ორთოგრაფიული შეცდომების კომპიუტერული შემოწმება

13.00-14.00 — შესვენება

14.00 **ჭ. აბაშიძე, თ. ჩხენკველი, ე. ჯგერენაია** — ციფრული კოლექციების განვითარება საქართველოს პარლამენტის ილია ჭავჭავაძის სახ. ეროვნულ ბიბლიოთეკაში

14.30 **ლ. ჩხაიძე** — ქართული ენის ელექტრონული გრამატიკული ლექსიკონისა და მართლმწერი სისტემის (სპელჩეკერის) ავანპროექტი

15.00 **ლ. ჩხაიძე, ი. ჯანაშია** — ლექსიკური პარადოქსული ქართული ტექსტები, როგორც ქართველურ ენათა ერთიანობის შესწავლის საშუალება

15.30-16.00 დისკუსია

ოთხშაბათი, 6 დეკემბერი

- 11.00 **ქ. ანთიძე, ნ. გულუა, დ. მელიქიშვილი, დ. მიშელაშვილი,
ლ. ნუკრაძე —** ინსტრუმენტული საშუალებების გამოყენება ქართული
ლი ტექსტების კომპიუტერული დამუშავებისათვის
- 11.30 **თ. ციციშვილი, ი. მოდებაძე —** ელექტრონული სამეცნიერო რეზე-
რირებული უურნალის ფორმის საკითხისათვის
- 12.00 **ი. მოდებაძე, თ. ციციშვილი —** ქართველოლოგიური მეცნიერების
საგანძურო (ელბიბლიოთეკა)
- 12.30-13.30 — შესვენება
- 13.30 **კ. ფხავაძე, ლ. ლეკიაშვილი, ვ. ფხავაძე —** მორფემების მათემატი-
კური ფუნქციებისა და წინადაღებებში სიტყვათშორისი ცარიელი
ადგილების პრობლემა ქართული ენის მიხედვით
- 14.00 **გ. ბაბუქაძია, ნ. ჩხაიძე —** „ვინდოუსის“ ქართულენოვანი ლოკალი-
ზაციის პროექტის განხორციელება
- 14.30 **ლ. ეზუგბაძა, თ. უთურგაიძე, რ. სხირტლაძე —** ქართული ენის
ზმნათა ფორმების ელექტრონული ვერსია
- 15.00-15.30 — დისკუსია

კონსტანტინე ფხავაძე, ვლადიმერ ლექიაშვილი, ვახტანგ ფხავაძე
ლოგიქისა და ენის გაერთიანებული ქართული ჯგუფი

მორფემების მათემატიკური ფუნქციებისა და
წინადადებებში სიტუგათშორისი ცარიელი ადგილების
პრობლემა ქართული ენის მიხედვით

კლასიკური მიდგომით, წინადადებებში სიტუგათშორისი ცარიელი ადგილები გაიგება კონკრეტული ცორმალურ ოპერაციად. ჩვენსული მიდგომით, ბუნებრივ ენობრივ სისტემაში სიტუგათშორისი ცარიელი ადგილების პრობლემა გაიაზრება იმ მათემატიკური ოპერაციის გამოცნობის პრობლემად, რომელიც ამ ენობრივ სისტემაში სიტუგათშორისი ცარიელი ადგილით მოგვეცება და რომელიც ამ სიტუგებით მოცემულ შინაარსებს ერთმანეთთან აზრობრივად აკავშირებს.

ბუნებრივი ქართული ენობრივი წესთვების პირდაპირი ფორმალურ-ლოგიკური აღწერის მიხედვით, სიტუგათშორისი ცარიელი ადგილები სისტემური თვალსაზრისით მიგვანიშნებენ იმ მათემატიკურ ოპერაციებზე, რომლებიც ზოგადი სიმრავლური და ლოგიკური წესების სახით ფუნქციონირებენ სისტემის ზედაპირულ და სილრმისეულ დონეებზე.

ჩვენ მიურ ჩატარებული ზოგადლინგვისტური კვლევების შედეგები გვარწმუნებენ, რომ ზემოაღნიშნული სახის მათემატიკურ ოპერაციებს შორის, წმინდა ლოგიკურ ოპერაციებთან ერთად, აუცილებლად უნდა იყოს ჩასმის ფორმალური ოპერაცია და სიმრავლიდან წერის ამორჩევისა და სიმრავლური თანაკვეთის მათემატიკური ოპერაციები. ეს საზოგადოდ, თუმცა, უკვე დაზუსტებით შეიძლება ითქვას, რომ ეს კი ის აუცილებელი მათემატიკური ოპერაციებია, რომელთა გარეშეც ქართული ენის მარტივი წინადადებები ვერ გაიგება. მაგალითისათვის განვიხილოთ წინადადება, რომელიც ჩვენი მიდგომის სადემონსტრაციოდ შემოგვთავაზა პ. პარტიმ:

(I) ჯონმა სთხოვა მარის ეთხოვა ბილისთვის მოეყვანა ძაღლი წევულებაზე.

ხაზგასმით აღვნიშნავთ, რომ ქართული ენის თანამედროვე ლოგიკური გრამატიკის მიხედვით (I) წინადადება მარტივი ანუ ერთი ლინგვისტური პრედიკატით აგებული წინადადებაა. კერძოდ, იგი აგებულია სთხოვა³ პ-ადგილიანი ლინგვისტური პრედიკატით, რომლის აღწერითი სახეა სთხოვა³ (Nm, Ns, No/Vn-ი). ეს იმას ნიშნავს, რომ სთხოვა³ პრედიკატით აგებულ მარტივ წინადადებას აქვს სთხოვა³ (Nm, Ns, No) ან სთხოვა³ (Nm, Ns, Vn) ფორმალური სახე, სადაც N (resp. V), ჩოქსეის აღნიშნების შესაბამისად, სახელურ (resp. ზმნურ) ფრაზაზე მიგვანიშნებს, ხოლო -მა, -ს და -ი მორფები შესაბამისი ბრუნვის ფორმების განმაზოგადებელი მატიკიზირებელი აღნიშვნებია. ის, რომ (I) წინადა

დებაში ჯონმა (resp. მარის) არის Nმა (resp. Nb) ტიპის მარტივი სამსჯელო სახელური ფრაზა, ხოლო ეთხოვა არის V-ი ტიპის მარტივი სამსჯელო ზმნური ფრაზა, განპირობებულია -მა, -ს და კ- მორფებით. იგივე მორფები განაპირობებენ იმას, რომ (I) წინადადების პირველი სამი სიტყვათშორისი ცარიელი ადგილით აზრობრივად რეალიზდება ფორმალური ჩასმის მათემატიკური ოპერაცია, რისი მეშვეობითაც სთხოვა³ პრედიკატის სამსჯელო ადგილები კომპლექტდება შესაბამისი სამსჯელო ფრაზებით, რაც (I) წინადადების შემდეგ არასრულ ფორმალურ სახეს იძლევა.

სთხოვა³(ჯონმა, მარის, ეთხოვა_ბილისთვის_მოეყვანა_მაღლი_წვეულებაზე)

ელასიკური მიღომით, ეთხოვა და მოეყვანა კავშირებითი კილოს ზმნებია. ჩვენეული მიღომით, ისინი არის სამსჯელო ანუ არაპრედიკატული ზმნური ფრაზები და ეს იმ უბრალო მიზეზის გამო, რომ ჩვენ მათი შინაარსით კი არ ვაფორმებთ ლოგიკურ პრედიკაციას, არამედ ამ ლოგიკურ პრედიკაციას სწორედ მათი მეშვეობით მოცემულ სამსჯელო შინაარსებზე ვაყალიბებთ. ამგვარად, ეთხოვა და მოეყვანა პ-ადგილიანი ფუნქციონალებია და ისინი შემდგენაირად აღიწერებიან: ეთხოვა³(Nb, Noხოვის, No/V-ი) და მოეყვანა³(Nb, No, Nზე). ამ მიღომით, (I) წინადადებაში ეთხოვა³ ფუნქციონალური ოპერატორის შემდეგ განთავსებული ორი ცარიელი ადგილით აზრობრივად რეალიზდება ფორმალური ჩასმის მათემატიკური ოპერაცია, რისი მეშვეობითაც ეთხოვა³ ლინგვისტური ფუნქციონალის შესაბამისი ადგილები ზუსტდება შესაბამისი საოპერატორო სახელური ფრაზებით, რაც საბოლოო ჯამში შემდეგ ფუნქციონალურ სახელურ ფრაზას იძლევა: ეთხოვა³(მარის=Nb, ბილისთვის, მოეყვანა_მერი_წვეულებაზე). ანალოგიურია ვითარება მოეყვანა³ ფუნქციონალის შემთხვევაშიც ამგვარად, სთხოვა³ (ჯონმა, მარის, ეთხოვა³(N1b, ბილისთვის, მოეყვანა³ (N2b, ბაღლი, წვეულებაზე)) (I) წინადადების სრული სტანდარტული სახეა და იგი (I) წინადადებისაგან მიიღება ცარიელი ადგილებით ნაგულისხმეული და მორფებით რეალიზებული ჩასმის ოპერაციების თანმიმდევრობითი განხორციელებით (აქ მარის=N1b, ბილი=N2b). ახლა განვიხილოთ შემდეგი წინადადება:

(II) კაცმა სთხოვა ქალს ეთხოვა მეზობლისთვის მოეყვანა მეზობელი წვეულებაზე.

ქართულში არა გვაქვს არტიკლები. (II) წინადადების ლოგიკური ანალიზი გვარწმუნებს, რომ მათ ფუნქციებს კომბინირებულად იღებენ თავიანთ თავზე -მა, -ს, -ისთვის, -ი და -ზე მორფები და სთხოვა³ (N1მა, N2ს, ეთხოვა(N2ს, N3ისთვის, მოეყვანა(N3ს, N4ი, N5ზე) ლინგვისტური პრედიკატული ფორმა. მართლაც, (II) წინადადებაში კაცმა სიტყვა შინაარსობრივად მიგვანიშნებს კაცთა სიმრავლიდან გამოყოფილ ერთ რომელიმე წევრზე. ამგვარად, (II) წინადადებაში -მა მორფება ახდენს სიმრავლიდან წევრის ამორჩევის სიმრავლური ოპერაციის რეალიზებას, რაც იმასაც ნიშავს, რომ კაცმა {კაცი} გამოსახულებით აღნიშნულია კველა კაცთა სიმრავლეზე განსახულებული არასაკუთრივი (ეგზისტენციალური) კონსტანტა. ანალოგიურად, ქალს (II) წინადა-

დებაში არის {ქალი} სიმრავლეზე განსაზღვრული არასაკუთრივი (ეგზისტენციალური) კონსტანტია, მეზობლისთვის და მეზობელი არის {მეზობელი} სიმრავლეზე განსაზღვრული არასაკუთრივი (ეგზისტენციალური) კონსტანტია, წვეულებაზე არის {წვეულება} სიმრავლეზე განსაზღვრული არასაკუთრივი (ეგზისტენციალური) კონსტანტია. ამგვარად, ქართული მორფემები ახდენენ ფორმალური ჩასმისა და სიმრავლიდან წევრის ამორჩევის მათემატიკური ოპერაციების რეალიზებას. ახლა განვიხილოთ შემდეგი წინადადება:

(III) შეყვარებულმა ქაცმა სთხოვა ქალს მოეყვანა მეზობელი წვეულებაზე.

აქ შეყვარებულმა_ქაცმა წმა ტიპის სახელური ფრაზაა. და თუ ქაცმა {ქაცმი} სიმრავლეზე განსაზღვრული არასაკუთრივი (ეგზისტენციალური) კონსტანტია, შეყვარებულმა_ქაცმა არის {შეყვარებული_ქაცმი}={შეყვარებული}ი\{ქაცმი} სიმრავლეზე განსაზღვრული არასაკუთრივი (ეგზისტენციალური) კონსტანტია. ამგვარად, ბუნებრივ ქართულენობრივ აზროვნებით წესთვებია A და N ტიპის სიტყვებით სიმრავლეები მოიცემა და, შესაბამისად, მათ შორის არსებული ცარიელი ადგილით აზრობრივად რეალიზება სიმრავლური თანაკვეთის ოპერაცია. ამგვარად, (III) წინადადების მათემატიკური თარგმანი იძლევა როგორც შეზღუდულ ეგზისტენციალურ კვანტორებიან ფორმას (იხ. (ა)), ისე შეზღუდულ არასაკუთრივ და ეგზისტენციალურ კონსტანტებიან ფორმებს (იხ. (ბ)და (გ)). ამასთან, როგორც შედეგი, ხაზგასმით აღსანიშნავია, რომ ეს „თარგმანები“ მექანიკურად იგება წინადადებებში სიტყვათშორისი ცარიელი ადგილებისა და გრამატიკული მორფემების მათემატიკური ფუნქციების გამოცნობით.

(ა) $\exists x \in \{\text{შეყვარებული}\} \cap \{\text{ქაცმი}\} \exists y \in \{\text{ქალი}\} \exists z \in \{\text{მეზობელი}\} \exists u \in \{\text{წვეულება}\} : \text{თხოვა}(x, y, \text{მოეყვანა}^3(y, z, u))$

(ბ) $a \in \{\text{შეყვარებული}\} \cap \{\text{ქაცმი}\} b \in \{\text{ქალი}\} c \in \{\text{მეზობელი}\} d \in \{\text{წვეულება}\} : \text{თხოვა}(a, b, \text{მოეყვანა}^3(b, c, d))$

(გ) $\text{თხოვა}(\exists x \in \{\text{შეყვარებული}\} \cap \{\text{ქაცმი}\}, \exists y \in \{\text{ქალი}\}, \text{მოეყვანა}^3(\exists y, \exists z \in \{\text{მეზობელი}\}, \exists u \in \{\text{წვეულება}\}))$

მანანა ჩაჩანიძე

არ. ჩიქობავას ენათმეცნიერების ინსტიტუტი

ზონისართის პრობლემა ქართული ენის სრულიად ფორმალიზებული მოდელის აგებისათვის

ზონისართის პრობლემა ქართულ ენათმეცნიერებაში საკმაოდ დატალურადაა შესწავლითი, როგორც საკუთრივ ქართული, ისე ქართვე-

The results of the performed work revealed the necessity for the solving of a number of the technological problems, their unification and their resolution foreseeing the modern requirements (firstly commercial).

K. Pkhakadze, V. Lekiaishvili, V. Phkakadze
United Georgian Group of Logic and language

MATHEMACAL FUNCTIONS OF MORPHO MARKERS AND THE PROBLEM OF EMPTY AMONG WORDS IN SENTENCES ACCORDING TO GEOGIAN NATURAL LANGUAGE SYSTEM

According to classical formal approaches the empty in sentences are understood as the operation of concatenation. We have understood the problem of the empty places among words in sentences as the task of recognition that mental (i.e. mathematical) operation is assumed in a NLS under any concret concatenation.

According to Direct Formal-Logical Description(DFLD) of the Georgian Natural Language System (GNLS) the empty places are giving us the implice understanding about these general mathematical operations which function in systematically as general set-theoretic or lingual-logic rules at the non deep and deep levels of the GNLS. We think, that generally, in the collection of these mathematical operations there must be a formal operation of substitution and the set-theoretic operations of choosing an element from the set and intersection. In any case, we can say strictly, that in the GNLS the above mentioned mathematical operations are implicitly assumed in simple sentences by mathemacal properties of morpho markers and empty places in it. For example:

(G1) ჯონ-მა ოხოვა მარი-ს ე-თხოვა ბილ-ისათვის მო-ე-ყვანა
დაღლ-ი წაეულება-ზე
jon-ma txova mari-s e-txova bil-isatvis mo-e-qvana
dzagl-i cveuleba-ze
John asked Mary to ask Bill to bring the
dog to the party

According to DFLD of GNLS we can say that any simple sentence in GNLS has a predicate constraction. This means that any simple sentence is constracted by one verbal-logical predicate which is characterized by its placenees (arity) and by the type of these places. For example, in G1 $\text{txova}^3(N\text{-ma}, N\text{-s}, N\text{-i}/V_{N,i})$ is described form of the 3-place lingual-logical predicate txova^3 . This means, that simple sentences constracted by predicate txova^3 have either $\text{txova}^3(N\text{-ma}, N\text{-s}, N\text{-i})$ or $\text{txova}^3(N\text{-ma}, N\text{-s}, V_{N,i})$ formal face, where, according to Chomsky's notation, N (resp. V) indicats nouns (resp. verbs) and -ma, -s and -i are morpho markers of the Georgian declension system. According to this general approach N-ma (resp. N-s) denotes any N-ma (resp. N-s) type term (i.e. N-ma (resp. N-s) type noun phrase) and in the context of $\text{txova}^3(N\text{-ma}, N\text{-s}, N\text{-i} / V_{N,i})$ it indicates about asking person (resp.

asked person). By analogy, $N-i$ (resp. V_{N-i}) denotes any $N-i$ (resp. V_{N-i}) type term (i.e. $N-i$ (resp. V_{N-i}) type noun phrase) and in this context it indicates the sense of asking.

In the (G1) **jonma** (resp. **maris**) is $N-ma$ (resp. $N-s$) type term (because of the $-ma$ (resp. $-s$) morpho marker) and **etxova** is the V_{N-i} type term (because of the $e-$ morpho marker). Due to these morpho markers the first three empty places in (G1) sentence gave us the mental understanding of the mathematical operation of the formal substitution in the general lingual-predicate idea $txova^3(N-ma, N-s, N-i / V_{N-i})$. After this mental operation we gave the (G2).

(G2) $txova^3(jon-ma, mari-s, e-txova \text{ bil-isatvis mo-e-qvana } dzagli \text{ cveuleba-ze})$

 asked³(Johan, Mary, to ask Bill to bring the dog at the party)

By classical approaches in GNLS ‘etxova’ and ‘moeqvana’ are verbs in the subjunctive mood. According to DFLD of GNLS the verbs in the conditional mood are not lingual-logical predicates because they never gave a sentence independently. To us ‘e-txova’ and ‘mo-e-qvana’ are V_{N-i} type terms (i.e. V_{N-i} type noun phrases) which are 3-phase lingual-functional ideas. By DFLD of GNLS they are described as

‘etxova ($N-s, N-istvis, N-i / V_{N-i}$)’ - to ask³($N, N, N / V_N$)’

and

‘moeqvana ($N-s, N-i, N-ze$)’ - to bring³(N, N, to_the_N))’

According to this general and natural function-arguments approach the next two empty places in (G1) sentence (i.e. the two empty places after ‘etxova³’) gives us the mental understanding of the mathematical operation of the formal substitution in the functional lingual-functional idea ‘etxova³($N-s, N-istvis, N-i / V_{N-i}$)’ and as a result of this substitution we receive us ‘etxova³(meri-s= $N-s$, bil-istvis, mo-e-qvana_dzagli_cveulebaze)’. The word ‘bilistvis’ as $N-istvis$ type term is substituted in the $N-istvis$ type place and the phrase ‘moeqvana_dzagli_cveulebaze’ as the V_{N-i} type term is substituted in the V_{N-i} type place. Into the $N-s$ type place of the ‘etxova³’ is assumed by syntactic-semantic substitution ‘maris= $N-s$ ’. By analogy, the two empty places in (G1) which are located after ‘moeqvana³’ give us the mental understanding of the mathematical operation of the formal substitution in the general lingual-functional idea ‘moeqvana³($N-s, N-i, N-ze$)’ and as a result of this substitution we get ‘moeqvana (bilis= $N-s$, dzagli, cveulebaze)’. The word ‘dzagli’ as $N-i$ type term is substituted into the $N-i$ type place and the word ‘cveulebaze’ as $N-ze$ type term is substituted into the $N-ze$ type place. Into the $N-s$ type place of the ‘moeqvana³’ is assumed the ‘bilis = $N-s$ ’ by the syntactic-semantic substitution.

Now, we can conclude, that the next expression which is the standard predicative form of the (G1) sentence is obtained from the (G1) only by substituting in the ‘ $txova^3(N-ma, N-s, V_{N-i})$ ’ $N-ma$, $N-s$ and V_{N-i} type terms.

$txova^3(jonma, maris, etxova^3(\text{meri-}N-s, \text{ bil-}istvis, \text{ moeqvana}^3(\text{bilis-}N-s, dzagli-i, cveuleba-ze)))$

 asked³(Johan, Mary, to ask³(Mary= N , Bill, to bring³(Bill= N , the dog, to_the_party)))

Now we will consider next sentence:

კაცმა თხოვა ქალს გთხოვა მეზობლისთვის მოეყვანა
 მეზობელი წევულებაზე
 kacma txova qals etxova mezoblistvis moeqvana mezobeli
 cveulebaze
 a man asked a woman to ask a neighbor to bring a
 neighbor to the party

In Georgian we have no articles, but in this example the morpho markers **-ma**, **-s**, **-istvis**, **-i**, and **-ze** together with morpho properties of the composed verbal-predicate form ‘txova³(N1-**ma**, N2-**s**, e-**txova**(N2-**s**, N3-**istvis**, mo-e-**qvana**(N3-**s**, N4-**i**, N5-**ze**) which is a general formal face of the above given sentence give us by the word ‘kac-**ma** (=a man)’ the understanding of one of the element from the set of all mans. So, we can say, that in GNLS by the ‘-ma’ morpho marker is realized the set-theoretical operation of the choice a element from the set. Because of this, we can say that ‘kac-**ma**’ is non-proper (i.e. existential) constant symbol defined on the set of all mans which is denoted as {kacma}. By analogy, in GNLS ‘qal-**s** (=a woman)’ is non-proper constans defined on the set {qals}, ‘mezobl-**istvis** (= a neighbor)’ is non-proper constans defined on the set {mezoblistvis}, ‘mezobl-**i** (= a neighbor)’ is non-proper constans defined on the set {mezobli}, ‘cveulebaze =(at the party)’ is non-proper constans defined on the set {cveulebaze}.

So, we can conclude: by morpho makcers of the Geogian declension system two mathematical operations are realized: the set-theoretical operation of the choice a element from the set and the mathematical operation of the formal substitution.

Now we will consider next sentence

შეუვრებულმა კაცმა თხოვა ქალს მოეყვანა მეზობელი
 წევულებაზე
 s'eqvarebul-ma kac-ma txova qal-s moeqvana mezobeli
 cveuleba-ze
 a loved man asked a woman to bring a neighbor at the
 party

Here, ‘s’eqvarebul-ma kac-ma’ is the N-**ma** type composed term (i. e. noun phrase), and if ‘kac-ma’ is non-proper constant defined on the set {kacma}, ‘s’eqvarebul-ma kac-ma’ is the same one but defined on the set {s’eqvarebulma_kacma} = {s’eqvarebul-ma} \cap {kac-ma}. So in GNLS A (adjectiv) and N (noun) type words gave us the understanding of sets and by the empty place between A and N type words in the sentences mentally is assumd the set-theoretic operation of the sets intersection. So, the next formula constracted by restricted quantifiers is the result of an automatical translation of the above sentence into mathematical language.

$$\exists x \in \{s'eqvarebulma\} \cap \{kacma\} \exists y \in \{qals\} \exists z \in \{mezobeli\} \exists u \in \{cveulebaze\}$$

$$: txova^3(x,y,moeqvana^3(y,z,u))$$

$$\exists x \in \{\text{loved}\} \cap \{\text{man}\} \exists y \in \{\text{woman}\} \exists z \in \{\text{neighbor}\} \exists u \in \{\text{party}\}: \text{asked}^3$$

$$(a_x, a_y, \text{to bring}^3(\text{the_}y, a_z, \text{at_}u))$$

Conclusion: the mathematical translation of the Georgian sentences can be make automatically according to mathematical properties of morpho markers and beetwen word's empty places.