

Reports of Enlarged Session of the  
Seminar of I.Vekua Institute  
of Applied Mathematics  
Vol. 19, N1, 2004

## **INSTRUMENTAL TOOL FOR COMPUTER MORPHOLOGICAL ANALYSIS OF SOME NATURAL LANGUAGES**

Antidze J., Mishelashvili D.

*I.Vekua Institute of Applied Mathematics  
Tbilisi State University*

Received in 15.05.04

The work has been done in order to automatize the computer morphological analysis of natural language. A specific formalism has been worked out, which simplifies the composition of a morphological analysis program.

The main feature of this formalism is as follow: the process of specific morphemes identification in a wordform is separated from the checking of that rules, which should be met by identified morphemes. The mentioned formalism implies the simplified recording both of the rules, that have to be checked and the ones that have been approved. Based on the given formalism a program product has been created, which by specific means of information transmission creates a program of morphological analysis. This program has been named a morphological analyser. By means of this program, the program for identification morphological categories of Georgian wordforms has been composed [1].

The program has not been created for a specific language and represents a generalized analyser. Although Georgian language has been used for demonstration of the program's possibility. It can be successfully used for other natural languages, as Russian, French, German, Hungarian, Arabic, Turkish etc.

To identify existing morphemes in a wordform, the concept of possible morpheme sequence has been introduced and on the basis of the concept, morphemes have been divided into classes. Certain wordforms might not have appropriate morphemes from certain classes. This complicates the process of identification of real morphemes in a wordform, because on the basis of morphemes omonimy alternate morphemes could be formed and they should be filtered later by placing certain constraints on morphemes. Such constraints represent logical expressions in which features and their values of morphemes participate. The features and values are determined by the linguist, which should consider the existing rules for a concrete language.

During the work on the program the following basic methods have been used: algorithm of nondeterministic search of morphemes in a wordform, feature structures with specific operation on them and the form of representation of constraints. A text file describing morphology of the natural language represents information for the program. The information in it is recorded by means of a special formalism. This formalism provides analyser both with the multitude of morphemes existing in a language and record in the rules the constraints that wordform morphemes have to comply with. Constraint

mechanism also enables determination of new information for a given wordform. Search mechanism, which divides a wordform into constituting morphemes, enables us with all of existing possibilities of a wordform division. Checking of constraints is conducted in parallel with the search process, which timely excludes all unwanted alternatives. Reshuffle of constraints is considered in a way that makes it possible to check the correctness of alternatives at the earliest stage. Constraints represent logical expression that is created by specific operations conducted on feature structures. In case logical expression receives a false value, search algorithm will reject the received version and will proceed searching from another version.

Formalism, which is used by the morphological analyser, is very useful. It has the number constructions that simplifies information recording. It has an installed processor that enables the usage of parametric macro insertions. The given formalism is first of all for linguists. With the analyser it is possible to record in more compact way morphological rules for natural languages. Future use of the program of morphological analyser enables linguist to verify correctness of recorded rules, so that they could be corrected if necessary.

Morphological analyser represents a program that can become a powerful tool in the hands of a linguist after she/he gets familiar with the formalism used in it.

The program is written by programming language C++ standard and uses STL standard library. It operates in UNIX and WINDOWS operational systems although it could be compiled in any other system as well which contains modern C++ compiler.

## REFERENCES

1. Antidze J., Mishelashvili D.. Recognition of Georgian Wordforms and their Morphological Categories by Computer, in the volume.