Compiler of Finite-State Automaton
for the Morphological Processor of the Georgian Language

Liana Lortkipanidze
liana.lortkipanidze@tsu.ge

Department of Computer Sciences, Ivane Javakhishvili Tbilisi State University, 13, University str.

Complete dictionaries of natural languages do not exist. It is also impossible to count an infinite set of numbers or to list all existing proper names. All languages undergo changes through the course of time. Their lexicons do too. All subsystems of a language have their expressions and a dictionary of a language can in no way include the complete lexicon of all of its individual varieties. Georgian has about seventeen dialects. Activities for creating the Georgian Dialect Corpus (GDC) are under way. A part of the technological procedures of the corpus activities is a case in point in the present paper.

We will dwell upon the compilation of a dialect morphological processor.

We will present a rather simple and, simultaneously, perfect technique which, by way of adapting of the already existing processor of a standard language, enables to compile a dialect processor. By means of the tools of our software, the system is trained for various dialects by applying known morpho-phonemic rules.

In order to verify the method, chose a corpus of Georgian dialects. Based on the morphological pattern of Standard Georgian, We adapted the morphological processor and afterwards attempted to lemmatize and surface annotate the dialect corpus.

Paragraph 2 of the present paper will discuss the system of the compilation of the morphological processor with respect to Georgian, paragraph 3 will deal with the technique of the adaptation of the standard language processor for dialect varieties, paragraph 4 will description Morphological Analysis and paragraph 5 will address the related work.