

Reports of Enlarged Sessions of the
Seminar of I. Vekua Institute
of Applied Mathematics
Volume 33, 2019

THE TRIAL VERSIONS OF THE NEW DEVELOPING TOOLS OF THE
GEORGIAN UNIVERSAL SMART CORPUS *

Konstantine Pkhakadze Shalva Malidze Merab Chikvinidze
Giorgi Chichua David Kurtskhalia Nodar Oqroshvili
Konstantine Demurchev

Abstract. In the paper the trial versions of the new developing tools of the Georgian Universal Smart Corpus are overviewed. These trial versions were developed on the basis of Pkhakadze-Chikvinidze’s and Pkhakadze-Chichua’s methods [1 - 2], which are elaborated on the basis of Pkhakadze’s Logical Grammar of Georgian language [3 - 4]. Besides, these trial systems are elaborated within the PHDF-18-1228 project “In the European Union with Georgian and Abkhazian Languages, i.e. the Doctoral Thesis - Elaboration of the New Developing Tools and Methods of the Georgian Smart Corpus and Improvement of Already Existing Ones” [5 - 6], which is financed by Shota Rustaveli National Science Foundation of Georgia. They are also elaborated within the PhD thesis “Methods and Tools for the Automatic Intellectual Classification of Georgian Texts” [7]. At the same time, these doctoral theses are sub-projects of the long term projects named “Technological Alphabet of the Georgian Language” (launched from 2012) and “Plan - Program for Complete Technological Support of the Abkhazian Language” (launched from 2015) of the Center for the Georgian Language Technology of Georgian Technical University and, accordingly, they are aimed at the defense of Georgian and Abkhazian languages from danger of digital extinction in the rapidly forthcoming digital age [8].

Keywords and phrases: Georgian universal smart corpus. Logical grammar of Georgian language.

AMS subject classification (2010): 68T50.

1 Introduction. Firstly, it must be mentioned, that the problem, at which the article is aimed, is the danger of the digital extinction of Georgian state languages - Georgian and Abkhazian. Thus, with this aim, the first experimental version of the Georgian Universal Smart Corpus (GUSC) [1 - 2] was constructed on the basis of Pkhakadze’s Logical Grammar of the Georgian Language (LGGL) [3 - 4] within FR/362/4-105/12 project “Foundations of LGGL and Its Application in Information Technology” (Supervisor K. Pkhakadze), which was financed by Shota Rustaveli National Science Foundation (SRNSF) in 2013-2015. In 2017, on the basis of this experimental version, within AR/122/4-105/14 project “One Step Toward Talking Georgian Self-Developing Corpus” (Supervisor K. Pkhakadze), which was financed by SRNSF in 2015-2017, first trial version

*The paper has been prepared with the financial support from the grant PHDF-18-1228 by Shota Rustaveli National Science Foundation.

of the GUSC was constructed [1]¹. Here we were based on Pkhakadze-Chikvinidze's and Pkhakadze-Chichua's methods [1], which were elaborated on the basis of Pkhakadze LGGL by the researches, carried out within DO/305/4-105/14 "In the European Union with the Georgian Language, i.e., the Doctoral Thesis - Georgian Grammar Checker (Analyzer)" (PhD student M. Chikvinidze; Scientific supervisor K. Pkhakadze) and DO/308/4-105/14 "In the European Union with the Georgian Language, i.e., the Doctoral Thesis - Georgian Speech Synthesis and Recognition" (PhD student G. Chichua; Scientific supervisor K. Pkhakadze) projects. These projects were financed by SRNSF in 2014-2016. After, the trial version of the GUSC was expanded and, now, the GUSC together with Georgian self-developing corpus contains the Abkhazian, Chechen, Kabardian, Lezgian and Mingrelian self-developing corpuses (<https://corpus.ge/>) [2].

Thus, in this stage of researches and, accordingly, in this paper shortly overviewed tools are aimed at the developing Georgian and Abkhazian self-developing corpuses, in other words, they are aimed to defend Georgian and Abkhazian languages - Georgian state languages from danger of the digital extinction in the rapidly forthcoming digital age [8].

2 The very short overview of the trial versions of the new developing tools of the Georgian universal smart corpus. It is overviewed here that trial versions of the new developing tools of the GUSC, namely, of the Georgian and Abkhazian self-developing corpuses, which were constructed:

2.1. Within PHDF-18-1228 project "In the European Union with Georgian and Abkhazian Languages, i.e. the Doctoral Thesis - Elaboration of the New Developing Tools and Methods of the Georgian Smart Corpus and Improvement of Already Existing Ones" (PhD Student Sh. Malidze, Scientific supervisor K. Pkhakadze). Thus, to improve the first trial-applied version of GUSC [1 - 2, 5 - 6, 8]:

2.1.1. Tools connecting the users' computers to corpus are constructed. - This result change qualitatively the corpus by equipping it with properties of the unified Georgian internet system.

2.1.2. Tools for constructing titrated speech corpuses for Georgian and Abkhazian languages are constructed. - This result gives an obligatory basis for the construction of higher quality speech processing systems for Georgian and Abkhazian languages.

2.1.3. On the basis of the first trial version of Abkhazian reader system the first trial version of Abkhazian voice-managed reader system is constructed. The system as subsystems contains the first trial version of the Abkhazian speech recognizer and Abkhazian speech synthesizer constructed based on the basis of syllables of Abkhazian language. - This result gives us substantial basis to build up interactive Abkhazian corpuses and technologies.

¹It should be emphasized that, in 2017, AR/122/4-105/14 project, because of its very important results, was announced as successful project of SRNSF (for more details see materials, which are located at the addresses <http://www.rustaveli.org.ge/en/Success-Stories/page/1988>; <https://geoanbani.com/other/pdfs/aboutus/2.pdf>).

2.1.4. The first improved trial-applied and trial versions of the Georgian Automatic Morpho-Syntactic-Semantic Annotators for Georgian N, A and V type words are constructed. - These results give us significant grounds to improve Georgian orthographic and grammatic text checkers.

2.2. Within PhD thesis “Methods and Tools for Automatic Intellectual Classification of Georgian Texts” (PhD student C. Demurchev; Scientific supervisor K. Pkhakadze). Thus, for improving the first trial-applied version of GUSC the first trial version of the system for automatic intellectual classification of Georgian texts is constructed. The classification system as subsystem contain the preliminary text processing tool and the text classification tool, which will be classified as text using parameters and functions collected by the first subsystem. - This result changes qualitatively the corpus by equipping it with properties of the unified Georgian internet system.

In addition to the above mentioned, within PhD thesis “Towards Georgian-Mathematical Automated Multilingual Semantic Translator” (PhD student N. Okroshiashvili; Supervisor K. Pkhakadze)², the tools for improving translating and intellectual abilities of GUSC are elaborated on basis of the methods of Pkhakadze-Chikvinidze and Pkhakadze-Chichua.

3 Conclusion. Considering the high-level threat of digital extinction in which the Georgian, even more the Abkhazian language is the elaboration of the trial tools, which is overviewed in the paper, should be assessed as one more important step toward defense of Georgian and Abkhazian languages from the danger of digital extinction in the rapidly forthcoming digital age.

R E F E R E N C E S

1. PKHAKADZE, K., CHIKVINIDZE, M., CHICHUA, G., KURTSKHALIA, D., MALIDZE, SH. The Georgian Intellectual Web - Corpus: Aims, Methods. Recommendation, Additional Publication of the Journal *Georgian Language and Logic*, (2017), 4-320.
2. PKHAKADZE, K., CHIKVINIDZE, M., CHICHUA, G., KURTSKHALIA, D., MALIDZE, SH. Georgian Universal Smart Corpus And In It Inbuilt Abkhazian, Chechen, Kabardian, Lezgian, And Mingrelian Self-Developing Corpuses - Results, Perspectives, *Rep. of Enlarged Sess. Semin. I. Vekua Inst. Appl. Math.*, **32** (2018), 55-58.
3. PKHAKADZE, K., CHIKVINIDZE, M., CHICHUA, G., KURTSKHALIA, D. A Logical Grammar of Georgian Language: Foundations and Applications - Third Part - Applications of The Logical Grammar of Georgian Language, Journal *Georgian Language and Logic*, **10** (2015-2016), 4-204.
4. PKHAKADZE, K. On the Linguistic Relation and Logical Declension in the Georgian Language, Journal *Georgian Language and Logic*, **1** (2005), 19-77.
5. PKHAKADZE, K., KURTSKHALIA, D., CHIKVINIDZE, M., CHICHUA, G., MALIDZE, SH. Toward The Abkhazian Talking Browser - Results and Perspectives. *Book of Abstract of the IX International Conference of Georgian Mathematical Union*, (2018), 187-188.

²This PhD project, as well as before mentioned PhD projects, is a subproject of long-term projects “Technological Alphabet of Georgian Language” and “Plan-Program for Complete technological support for Abkhazian Language”.

6. PKHAKADZE, K., CHICHUA, G., CHIKVINIDZE, M., KURTSKHALIA, D., MALIDZE, SH. Toward the Computer and Internet Versions of the Abkhazian Voicemanaged Reader System - Results and Perspectives. *Book of Abstract of the IX International Conference of Georgian Mathematical Union*, (2018), 183-184.
7. DEMURCHEV, K., PKHAKADZE K. The General Overview of the Aims, Tasks and Methods of the Doctoral Thesis – “Methods and Tools for Automatic Intellectual Classification of Georgian Texts”, *Book of Abstract of the IX International Conference of Georgian Mathematical Union*, (2018), 105-106.
8. PKHAKADZE, K., CHIKVINIDZE, M., CHICHUA, G., KURTSKHALIA, D., MALIDZE, SH. The Open Letter to the Georgian Parliament, Government, National Academy of Sciences and Georgian and Abkhazian Society i.e. Key Principles of the Unified program of Complete technological supporting of the official language of Georgia (Georgian, Abkhazian) i.e. In the Future Cultural World with the Completely Supported Georgian and Abkhazian Languages, *Journal Georgian Language and Logic*, **11** (2017-2018), 21-164.

Received 28.05.2019; revised 20.11.2019; accepted 15.12.2019.

Author(s) address(es):

Konstantine Pkhakadze, Shalva Malidze, Merab Chikvinidze, Giorgi Chichua,
David Kurtskhalia, Nodar Oqroshvili, Konstantine Demurchev
Center for Georgian Language Technology, Georgian Technical University
77, Kostava Str., 0160, Tbilisi, Georgia
E-mail: gllc.ge@gmail.com