

THE TECHNOLOGICAL ALPHABET OF THE GEORGIAN LANGUAGE -
AIMS, METHODS, RESULTS

Pkhakadze K., Chikvinidze M., Chichua G., Maskharashvili A.

Abstract. In 2012, in the Center for the Georgian Language Technology (GLTC) at the Georgian Technical University (GTU), there was started a project “Georgian Technological Alphabet (hereafter GTA)”, which is lead by K.Pkhakadze and which is elaborated on the basis of further development of the state priority program “Free and Complete Inclusion of A Computer in Natural Georgian Language System”, which, also, was lead by K.Pkhakadze. In particular, in 2012, at the GTU, there were started doctoral theses “Georgian Speech Synthesis and Recognition” and “Georgian Grammar Checker (Analyzer)” (doctoral students: G.Chichua and M.Chikvinidze, supervisor: K.Pkhakadze - the full professor of the Department of Artificial Intelligence at GTU), which have fundamental importance in the process of creating the GTA. Furthermore, in the GLTC, on 25th of April 2013, there was started a project “Foundations of the Logical Grammar of the Georgian Language and its Application in Information Technology”, which is funded by Rustaveli Foundation¹, and on 1st of October 2013, there was started another project: “The Extensible Internet Versions of the Basic Systems for Creating the GTA” (the leader of both of the projects: the director of the GLTC K.Pkhakadze). The researches, which are being pursued in the confines of the above-mentioned projects and doctoral theses, are considered to have fundamental importance for the preparatory stage of the project GTA. Thus, in the paper, we overview the aims, methods and results of our above-underlined researches aimed at the creation of GTA. Also, we pay special attention to the alarming lagging condition of the Georgian language in which the Georgian language is in the digital age.

Keywords and phrases: The logical grammar of the Georgian language, The technological alphabet of the Georgian language, The trial version of the Georgian self-developing intellectual corpus, The Georgian language in the digital age.

AMS subject classification: 03B65, 68T50, 68Q55, 91F20.

1. The aims, methods and results of creating Georgian technological alphabet. To build GTA alphabet means to fully equip a computer with the intellectual mathematical theory of the Georgian thinking and communication systems, which are naturally given by Georgian spoken and written languages. In other words, this means the creation in the Georgian language freely interacting intellectual computer systems - Georgian talkie intellectual computer systems [7]. Thus, to build GTA means to

*This publication was supported by Shota Rustaveli National Science Foundation grant N 31/70 and Georgian Technical University grant N 048-13.

¹A referee of the project outlined following: “If successful, the publication of the ‘Foundations of Logical Grammar of Georgian Language’ may be considered an historical contribution to Georgian linguistics and information technology”. - Indeed, already existing preliminary version of the logical grammar of the Georgian language is historically first mathematical theory of the Georgian language.

build Georgian lingual computer brain, ear and throat, which will allow any Georgian speaking user to interact freely i.e. without having any specific computer knowledge with the computers that know the Georgian language or with any type of technological device that knows Georgian language, which devices for today are called smart things and whose number and kind are increasing day by day in parallel with further perfection and realization of the idea of smart houses. Besides this, in the aims of building up GTA, there is also an extension of the Georgian talkie intellectual computer system with the translation abilities as are from speech to speech and from text to text, which besides above-mentioned, will give abilities to a Georgian user to interact with foreign users of the Internet without knowing their foreign languages and also to find in the Internet the information written in any language. It is clear, that in the near future, without the GTA a Georgian user will be almost completely restricted in his/her working activity by means of the Georgian language. It is also clear that without the GTA, the Georgian pre-school children will not be able to play with Georgian speaking smart toys, which help children to develop their intellectual abilities more; Moreover, the Georgian school students will not be able to get involved in educational system equipped with smart educational technologies. Thus, the final aim of the project GTA is the fully technological elaboration of the Georgian language, without which in the forthcoming digital age, it will be impossible to maintain the Georgian language as a such language by means of which cultural processes go on as it is for today. This circumstance makes very clear very high national actuality of the project.

As it was already mentioned, toward creation of the GTA, our main theoretical basement is the intellectual mathematical theory of the Georgian thinking and communication systems. According to our approach, a mathematical theory of a communication system of any language should be based on the mathematical theory of the thinking system of this language, because for us lingual thinking, which is a process going on in one linguistic individual, is not characterized by those polysemy problems that arise during linguistic communication [4],[6], i.e. interaction between two linguistic individuals. Thus, for us, the main methodological line of elaboration of the GTA goes through the creation of the mathematical theory of the Georgian thinking system - the logical grammar of the Georgian language [4],[6]. In the process of systemization of the fundamentals of the Georgian logical grammar, which is the main aim of our current researches, in accordance with Georgian Language's theses [5], we base on sufficiently proven possibility of describing the Georgian language [3], [4] on the basis of Sh.Pkhakadze's sufficiently general mathematical theory \mathcal{T} [1]; we also base on the view that the Georgian language and respectively the Georgian thinking system is a result of the step-by-step extension of the primary mathematical language and theory, which exist in every human as universal, i.e. inborn data, with Sh.Pkhakadze's type of contracting symbols [5]. Thus, together with Chomsky-Montague's general logical-linguistic views [2], we mainly base on Sh.Pkhakadze's Notation Theory [1], K.Pkhakadze's methods of direct formal-logical, i.e. natural mathematical description of Georgian language [3], [4], [5], and, also, on the program of foundation of the Georgian thinking system as lexically and semantically developable system [6]. Hence, our activity is directed towards the representation of the Georgian thinking system as a rule-based system, and despite the fact that we base on Montague's general set-theoretic

semantic approaches, our semantic approaches are closer to Pascal's and Leibniz's ones rather to those ones which are dictated by Frege-Montague's principle of compositionality. Here we mean Pascal's view on the necessary existence of the words that have naturally inborn semantics understandings, and Leibniz's view on the linguistically given and logically developable knowledge existing in humans. However, in the process of creating GTA i.e. Georgian lingual computer brain, ear and throat, we also use the statistical methods based on the Georgian language corpus. Namely, we already use a corpus that is created by us using very optimal ways, which is a trial version of the Georgian self-developing intellectual corpus. This corpus, together with other unique systems for Georgian created by us, can be found at <http://geoanbani.com/>

Namely, at <http://geoanbani.com/> one can find our experimental systems, which have no other Georgian analogs. They are: 1. Georgian developable intellectual corpus; 2. The Georgian orthographic spell-checker based on our Georgian language corpus; 3. The Georgian text reader; 4. The Georgian speech recognizer working on base of teaching and self-teaching principle; 6. The Georgian extension of Google translator; 7. The Georgian syntax analyzer 8. The Georgian-English and Georgian-German rule based two-way translators. Together with these already listed systems in GLTC, with the aims of the elaboration of the GTA we have already created trial versions of such Georgian systems as are: 1. Georgian logical analyzer; 2. Georgian-Mathematical two-way translator; 3. Multilingual i.e. Georgian-English-German rule based two-way translator; 4. The first experimental Georgian speech to speech translator system. - The Internet versions of these systems will be placed at the address <http://geoanbani.com/> in the near future.

2. The Georgian language is at especially high risk. Above, we shortly presented the trial systems elaborated by us on the basis of the logical grammar of the Georgian language in the GLTC. Also, it was underlined that these systems, besides the orthographic checker, are unique in the sense that outside of the GLTC, unfortunately, there are no not only comparable, but even any kind of alternative systems. On the other hand, these systems, which are unique from the local Georgian point of view and, accordingly, which are the highest quality experimental systems for Georgian, can be hardly compared to the same kind of systems which exist for well studied and processed languages. Here together with the information that is available to us, we are also mainly based on the evaluations of the technological support of European languages, which were prepared for 32 European languages by the white papers series of Meta-Net (<http://www.meta-net.eu/whitepapers/overview>). In this way, the fact that European languages, which are much better studied and elaborated languages than Georgian, are considered to be under the risk in the forthcoming digital age, i.e. in the forthcoming epoch of technologically fully elaborated languages, makes it very clear that for today the Georgian language is at especially high risk. This makes absolutely clear the urgent necessity of the high governmental support for defending and maintaining the future of the Georgian language.

REFERENCES

1. Pkhakadze Sh. Some Problems of The Notation Theory. (Russian) *Tbilisi University Press, Tbilisi*, 1977.
2. Handbook of Logic & Language, *Editors J. van Benthem and A. ter Miulen, North-Holland*, 1997.
3. Pkhakadze K., Chankvetadze G., Tibua L., Ivanishvili M., Lekiasvili, L., Soselia E. About the main ideas of the direct formal-logical description of the Georgian natural language system through the examples. (Georgian) *Proceedings of the V Tbilisi Symposium on Language, Logic and Computation, ILLC, University of Amsterdam, Tbilisi State University*, (2003), 129-137.
4. Pkhakadze K. About logical declination and lingual relations in Georgian. (Georgian) *Journal of Georgian language and logic*, **1** (2005), 19-77.
5. Pkhakadze K., Abzianidze L., Maskharashvili A. Georgian language's theses. *Semin. I. Vekua Inst. Appl. Math. Rep.*, **34** (2008), 96-107.
6. Pkhakadze K., Chichua G., Chikvinidze M., Maskharashvili A. Methods, and results of the logical grammar of the Georgian language. *Rep. Enlarged Sess. Semin. I. Vekua Appl. Math.*, **26**, (2012), 58-64.
7. Pkhakadze K. The Technological Alphabet of The Georgian Language The One of The Most Important Georgian Challenge of XXI Century. (Georgian) *Journal/Editor, The Works of The Parliament Conference "The Georgian Language - The Challenge of the 21st Century"*, 2013.

Received 12.05.2013; revised 17.09.2013; accepted 24.11.2013.

Authors' addresses:

K. Pkhakadze, M. Chikvinidze, G. Chichua
Center for the Georgian Language Technology
Georgian Technical University
77, Kostava Str., Tbilisi 0175
Georgia
E-mail: gllc.ge@gmail.com

A. Maskharashvili
Center for the Georgian Language Technology
Georgian Technical University
77, Kostava Str., Tbilisi 0175
INRIA-LORIA & University of Lorraine, Nancy
LORIA/INRIA - 615, rue du Jardin Botanique, 54500, Villers-lès-Nancy, France
E-mail: gllc.ge@gmail.com