საქართველოს მეცნიერებათა ეროვნული აკადემიის მოამბე, ტ. 14, №3, 2020
BULLETIN OF THE GEORGIAN NATIONAL ACADEMY OF SCIENCES, vol. 14, no. 3, 2020

Informatics

In the European Union with Georgian and Abkhazian Languages – Aims and Problems of Complete Technology Support of Georgian and Abkhazian Languages

Konstantine Pkhakadze[*], Merab Chikvinidze[*], Giorgi Chichua[*], Shalva Malidze[*], David Kurtskhalia[*], Constantine Demurchev[*], Nodar Oqroshiashvili[*]

[*]*Center for Georgian Language Technology, Georgian Technical University, Tbilisi, Georgia*

(Presented by Academy Member Archil Prangishvili)

On May 17, 2019, a report "In the European Union with Georgian and Abkhazian languages - Aims and Problems of Complete Technology Support of Georgian and Abkhazian Languages" was presented at the Center for Innovation and High Technologies, Georgian National Academy of Sciences. The present paper is a short publication version of the above-mentioned report, where the aims and problems of complete technology support for Georgian and Abkhazian languages are considered and reality of high level danger of digital extinction of Georgian and Abkhazian languages is proved. The recommendations supported by protocol #53 of 17 May 2019 of the Center for Innovation and High Technologies, Georgian National Academy of Sciences, are presented and the necessity of taking them into consideration is proved. *© 2020 Bull. Georg. Natl. Acad. Sci.*

Georgian and Abkhazian technological alphabets, challenges of the digital age

The aim of complete technology support of Georgian and Abkhazian languages implies to build Georgian and Abkhazian technological alphabets, i.e. computer systems completely knowing Georgian and Abkhazian languages. In other words, this implies to reach perfect solutions of the tasks of constructing artificial intelligence systems for Georgian and Abkhazian languages. These tasks are the problems of top difficulty that belong to the interdisciplinary scientific area of the logic, language, and artificial intelligence. At the same time, they are especially highly important for cultural future of Georgian and Abkhazian languages [1-4]. It is clear that only in the case of finding a perfect solution for them, it will be possible to defend cultural future of Georgian and Abkhazian languages in rapidly forthcoming digital age and, accordingly, to enter the European union, and generally, in the future digital world with technologically completely supported Georgian and Abkhazian languages [1-5].

Problems of complete technology support for Georgian and Abkhazian languages are mainly conditioned, first, by the difficulties of achieving the abovementioned aims, in other words, by the difficulties of constructing technological alphabets of Georgian and Abkhazian languages, second, by the absence of a unified Georgian research center, which will be targeted at solution of these very difficult and important tasks, and, third, by low quality of the language technology support for Georgian and Abkhazian languages [1-6].

On 20 September, 2012, on the basis of two-year study "Europe's Languages in the Digital Age", which was pursued in 2010-2012 years, meta-net published a very alarming press-release [7] according to which "a new study by Europe's leading Language Technology experts warns" that "most European languages are unlikely to survive in the digital age". This two-years "study, prepared by more than 200 experts and documented in 30 volumes of the meta-net white paper series, assessed language technology support for each language in four different areas: automatic translation, speech interaction, text analysis and the availability of language resources. A total of 21 of the 30 languages (70%) were placed in the lowest category, "support is weak or non-existent" for at least one area by the experts. Several languages, for example, Icelandic, Latvian, Lithuanian and Maltese, receive this lowest score in all four areas. On the other end of the spectrum, while no language was considered to have "excellent support", only English was assessed as having "good support", followed by languages such as Dutch, French, German, Italian and Spanish with "moderate support". Languages such as Basque, Bulgarian, Catalan, Greek, Hungarian and Polish exhibit "fragmentary support", placing them also in the set of high-risk languages."

Based on the foregoing, we rated "excellent support" at 5 points, "good support" at 4 points, "moderate support" at 3 points, "fragmentary support" at 2 points, "weak or no support" at 1 point and, then, on the basia of Table 7 from [8], we generated a Table given below. In this Table we assessed language technology support for Georgian and Abkhazian languages on the basia of results of the long-term projects "Technological Alphabet of Georgian Language" and "Plan-Program for Complete Technological Support of the Abkhazian Language" [1, 3, 5], which are being processed in the center for Georgian language technology of Georgian Technical University.

Thus, all these make it clear that European languages, which summary points are less or equal to 9 points (see Table), i.e. Hungarian, Polish, Czech, Galician, Catalan, Portuguese, Swedish, Romanian, Slovenian, Slovak, Greek, Finnish, Danish, Bulgarian, Basque, Norwegian, Estonian, Serbian, Irish Croatian, Icelandic, Latvian, Lithuanian, Maltese, Welsh languages are appreciated as languages under danger of digital extinction. This fact together with the fact, that language technology support for Georgian and Abkhazian languages is less than technology support for any aforementioned European languages, clearly shows a reality of a high-level danger of digital extinction of Georgian and Abkhazian languages in the rapidly forthcoming digital age.

The fact that language technology support for the Georgian and Abkhazian languages is really much less than technology support for European languages is also proved by below listed language technology systems, which are created within the projects "One More Step Towards Georgian Talking Self-Developing Intellectual Corpus" (in 2017 Shota Rustaveli National Science Foundation of Georgia announced this AR_122/4-105/14 project as a successful project of the year, for more details see web-page of the fund) and "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" (this PHDF-18-1228

project was funded by Shota Rustaveli National Science Foundation of Georgia in 2018-2019 years) according to Pkhakadze's, Chikvinidze's, Chichua's, and Malidze's new methods [5-6] elaborated on the basis of Logical Grammar of Georgian Language [9].

technological support of Abkhaz Language" [1-3], are based on the projects "Foundations of Logical Grammar of Georgian Language and its Application in Information Technology", "In the European Union with the Georgian Language, i.e. Doctoral Thesis – Georgian Grammar Checker

Table. State of language technology support for Georgian, Abkhazian and 31 European languages in four different areas in points and percents

Language	Speech interaction	Automatic translation	Text analysis	Language resources	Summary points/percents
1. English	4	4	4	4	16 = 80%
2. French	3	3	3	3	12 = 60%
3. Spain	3	3	3	3	12 = 60%
4. German	3	2	3	3	11= 55%
5. Dutch	3	2	3	3	11 = 55%
6. Italian	3	2	3	3	11 = 55%
7. Hungarian	2	2	2	3	9 = 45%
8. Polish	2	2	2	3	9 = 45%
9. Czech	3	1	2	3	9 = 45%
10. Galician	2	1	3	3	9 = 45%
11. Catalan	2	2	2	2	8 = 40%
12. Portuguese	3	1	2	2	8 = 40%
13. Swedish	2	1	2	3	8 = 40%
14. Romanian	1	2	2	2	7 = 35%
15. Slovenian	2	1	2	2	7 = 35%
16. Slovak	2	1	2	2	7 = 35%
17. Greek	2	1	2	2	7 = 40%
18. Finnish	2	1	2	2	7 = 40%
19. Danish	2	1	2	2	7 = 35%
20. Bulgarian	2	1	2	2	7 = 35%
21. Basque	2	1	2	2	7 = 35%
22. Norwegian	2	1	2	2	7 = 35%
23. Estonian	2	1	1	2	6 = 30%
24. Serbian	2	1	1	2	6 = 30%
25. Irish	2	1	1	1	5 = 25%
26. Croatian	1	1	1	2	5 = 25%
27. Icelandic	1	1	1	1	4 = 20%
28. Latvian	1	1	1	1	4 = 20%
29. Lithuanian	1	1	1	1	4 = 20%
30. Maltese	1	1	1	1	4 = 20%
31. Welsh	1	1	1	1	4 = 20%
32. Georgian	0.75	0.75	0.75	0.75	3 = 15%
33. Abkhazian	0.25	0.25	0.25	0.25	1 = 5%

At the same time, abovementioned projects, which are important subprojects of long-term projects "Technological Alphabet of Georgian Language" and "Plan-Program for Complete

(Analyzer)", "In the European Union with the Georgian Language, i.e. Doctoral Thesis – Georgian Speech Synthesis and Recognition" [9], which, in turn, are fundamental subprojects of the

long-term project "Technological Alphabet of Georgian Language" funded by Shota Rustaveli National Science Foundation of Georgia. Also, we underline that most of the below presented systems are unique in the sense that no other similar system of Georgian and Abkhazian languages exists today [3, 5-6]:

1. In the area of speech interaction. Trial-applied system of Georgian spoken browser; trial-applied (trial) voice managed systems of the Georgian (Abkhazian) readers and listeners; trial-applied (trial) systems of Georgian (Abkhazian) assistants for speech disorder persons; trial systems of Georgian adapted Internet, Wikipedia, and Computer.

2. In the area of automatic translation. Trial system of Georgian-Mathematical translator; trial system of Georgian-English-German semantic translator; trial-applied internet (mobile) hybrid systems of Georgian multilingual voice lexicons; trial internet (mobile) hybrid systems of Georgian multilingual spoken assistant; trial system of Georgian extension of Google translator; trial internet (mobile) systems of multilingual textual and voice messages between Georgian smart papers;

3. In the area of text analysis. Systems for automatic processing Georgian (Abkhazian) texts and web-sites; trial and trial-applied systems of taggers, descriptors and generators for Georgian words of V, N and A type; trial-applied (trial) system of Georgian (Abkhazian) self-developing orthograph checker; trial system of Georgian self-developing syntax checker; trial systems of Georgian text analyzer and question answerer; trial systems for automatic generating and testing of Georgian logical tasks and analogies, trial system of Georgian smart paper; trial-applied system of Georgian texts classification.

4. In the area of language resources. Self-developing Georgian (Abkhazian) Trial-applied (trial) multilingual and multimodal corpuses; trial system of analyzer, segmentator and generator for Georgian (Abkhazian) titrated speech data; trial-applied (trial) voice inbuilt tools for the Georgian (Abkhazian) reader-listener; trial-applied (trial) systems of Georgian (Abkhazian) printed and scanned texts recognition.

On the basis of the sufficient strict observations made above, we can say that, the language technology support for Georgian (Abkhazian) language is 15% (5%), when support for English language is 80%. This is conditioned by the fact that if we have only the trial-applied and trial systems for Georgian and Abkhazian languages, the same type of applied and trial-applied systems with sufficiently high quality are already prepared for English as well as for many other European languages.

These means, that when only 20% is left to achieve the aim of complete technology support for English language, for Georgian language lacks 85%. Thus, if Georgian language is under the high-level danger of digital extinction, Abkhazian language, which lacks 95% to achieve the aim of complete technology support, is really under the especially high-level danger of digital extinction.

All the above said shows that, for today, Georgian, even more Abkhazian language is alarmingly lagging in the sense of the language technology support compared with technologically advanced and, even not advanced European languages. This, together with the fact that in recent years European languages are doing quick steps for providing excellent language technology support, i.e. for constructing technological alphabets for their languages, shows the necessity of further rapid development of the already ongoing local processes for constructing technological alphabets of Georgian and Abkhazian languages, the construction of which is a vital challenges of the digital age for Georgian and Abkhazian languages, i.e. for the Georgian state [1-6].

Thus, all above said clearly proves the reality of high-level danger of digital extinction of Georgian and Abkhazian languages in the rapidly forthcoming

digital age and, it is also clear that this must be taken into consideration by the Georgian state.

Conclusion

Taking into consideration the objectives of defense of the state languages of Georgia – Georgian and Abkhazian languages – from danger of digital extinction the below given recommendations were fully approved by the Protocol #53 of 17 May 2019 of the Center for Innovation and High Technologies, Georgian National Academy of Sciences. These recommendations prove that it is necessary: 1. To form the "Research Institute for Technological Development and Cultural Defense of Georgian Official Languages" under the joint aegis of the Georgian National Academy of Sciences and the State Language Department. 2. To define "Technological development and cultural defense of Georgian state languages" as one of the main priorities of the Shota Rustaveli National Science Foundation of Georgia. 3. To hold the conference "Georgian State Languages in the Digital Age: Cultural and Technological Aspects" twice a year (on April 14 – Georgian language day, and on October 27 – Abkhaz language day) under the joint aegis of Georgian National Academy of Sciences and State Language Department, that will allow constantly to observe current dynamics and results of processes with the aims of the cultural defense and technological development of the Georgian and Abkhazian languages.

It is clear that by taking the above listed recommendations into account, first, it becomes clear that in the forthcoming digital age Georgian state really takes care of protecting Georgian and Abkhazian languages from the rapidly increasing danger of the digital extinction, and, second, it will also become clear to Abkhazians, who live in occupied Abkhazia, that it is vitally important for the Georgian state to preserve Abkhazian language and identity.

Thus, it is also clear that all the above written will be a very important step in the direction of reconstruction of the present broken bridge in the Georgian-Abkhazian relations.

The work was completed by the grant support of the PHDF-18-1228 project of Shota Rustaveli National Science Foundation of Georgia.

ინფორმატიკა

ქართული და აფხაზური ენებით ევროკავშირში ანუ ქართული და აფხაზური ენების სრული ტექნოლოგიური უზრუნველყოფის მიზნები და პრობლემები

კ. ფხაკაძე*, მ. ჩიქვინიძე*, გ. ჩიჩუა*, შ. მალიძე*, კ. დემურჩევი*, დ. კურცხალია*, ნ. ოქროშიაშვილი*

* *საქართველოს ტექნიკური უნივერსიტეტი, ქართული ენის ტექნოლოგიების ცენტრი, თბილისი, საქართველო*

(წარმოდგენილია აკადემიის წევრის ა. ფრანგიშვილის მიერ)

2019 წლის 17 მაისს საქართველოს მეცნიერებათა ეროვნული აკადემიის ინოვაციებისა და მაღალი ტექნოლოგიების ცენტრში გაკეთდა მოხსენება „ქართული და აფხაზური ენებით ევროკავშირში ანუ ქართული და აფხაზური ენების სრული ტექნოლოგიური უზრუნველყოფის მიზნები და პრობლემები“. ამგვარად, ნაშრომში, რომელიც ამ მოხსენების მოკლე საპუბლიკაციო ვერსიაა, დასაბუთებულია ციფრული კვდომის იმ მაღალი საფეხურის საფრთხის რეალობა, რომლის წინაშეცაა ქართული და აფხაზური ენები; მოკლედაა მიმოხილული ქართული და აფხაზური ენების სრული ტექნოლოგიური უზრუნველყოფის მიზნები და პრობლემები; წარმოდგენილია აკადემიის ინოვაციებისა და მაღალი ტექნოლოგიების ცენტრის 2019 წლის 17 მაისის #53 ოქმით სრულად მხარდაჭერილი რეკომენდაციები და დასაბუთებულია მათი გათვალისწინების აუცილებლობა.

REFERENCES

1. Pkhakadze K. (2013) The technological alphabet of the Georgian language – the one of the most important Georgian challenges of the XXI century. *The works of Parliament Conference "Georgian Language – Challenges of 21th Century",* 98-105.
2. Pkhakadze K., Chikvinidze M., Chichua G., and Kurtskhalia D. (2016) Take care on the Georgian language. *Proceedings Georg. Natl. Acad. Sci. – series of History, Archeology, Ethnology and Art History,* **1:** 149-158.
3. Pkhakadze K., Chikvinidze M., Chichua G., Kurtskhalia D., Malidze Sh. (2018) 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. *Georgian Language and Logic,* **11:** 121 – 164.
4. Pkhakadze K. (2009) For national security conception. *Georgian University*, **49:** 9-11.
5. Pkhakadze K., Chikvinidze M., Chichua M., Kurtskhalia D., Malidze Sh., Beriashvili I. (2017) The Georgian intellectual web – corpus: aims, methods, recommendation. *Additional Edition of Journal Georgian Language and Logic,* 1-320.
6. Pkhakadze K., Chikvinidze M., Chichua G., Kurtskhalia D., Malidze Sh. (2018) Georgian universal smart corpus and in it inbuilt Abkhazian, Chechen, Kabardian, Lezgian and Mingrelian self-developing corpuses – results, perspectives. *Reports of Enlarged Session of the Seminar of VIAM,* **32:** 55-58.
7. Burchardt A., Rehm G., and Uszkoreit H. (2012) Strategic research agenda for multilingual Europe 2020. *Presented by the Meta-Net Technology Council,* 1-98.
8. Uszkoreit H., Rehm G. (2012) META-NET White Paper Series: Press Release "At Least 21 European Languages in Danger of Digital Extinction – Good News and Bad News on the European Day of Languages". *Presented by the Meta-Net Technology Council.*
9. Pkhakadze K., Chikvinidze M., Chichua G., Kurtskhalia D. (2015-2016) A Logical Grammar of the Georgian language: foundations and applications – third part – applications of the Logical Grammar of the Georgian language. *Georgian Language and Logic,* **10:** 4-204.

Received December, 2019