

PHILOLOGY

THE GOAL AND ISSUES OF THE FORMAL DESCRIPTION OF ARMENIAN VOCABULARY AND CREATION OF THE ELECTRONIC DATABASE**PhD Sargsyan M. A.*

Armenia, Yerevan, NAS RA Language Institute/ NAS RA ISEC

DOI: https://doi.org/10.31435/rsglobal_wos/31012019/6319**ARTICLE INFO****Received:** 20 November 2018**Accepted:** 14 January 2019**Published:** 31 January 2019**KEYWORDS**formal description,
Armenian language,
language vocabulary,
electronic database,
linguistic data**ABSTRACT**

The article is devoted to the goals and issues of formal description of Armenian word formation. Although recently formal description of the language has become more practical, it should be noted that the full formal description of the Armenian language has not been done yet. The author of the article states that the formal description of the Armenian vocabulary will enable to reveal as the regular structures, as well as the variative forms, deflections, and irregularities with frequency data with their automatic analysis and the possibility of derivation. It has an informative value, it can clarify and explain many problems in Armenian linguistics, and also can give much material for the further researches of Armenian grammar.

Citation: Sargsyan M. A. (2019) The Goal and Issues of the Formal Description of Armenian Vocabulary and Creation of the Electronic Database. *International Academy Journal Web of Scholar*. 1(31), Vol.2. doi: 10.31435/rsglobal_wos/31012019/6319

Copyright: © 2019 Sargsyan M. A. This is an open-access article distributed under the terms of the **Creative Commons Attribution License (CC BY)**. The use, distribution or reproduction in other forums is permitted, provided the original author(s) or licensor are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Introduction. Although recently formal description of the language has become more practical, it should be noted that the full formal description of the Armenian language has not been done yet.

However, the fact that certain attempts have been made is undeniable. Particularly, dissertations [1] and some scientific articles [6] are dedicated to the study of formal description of Armenian language, but these have just a theoretical value and have not been used practically.

An attempt has recently been made to introduce Armenian in the UNL environment. A working group was formed, some principles have been developed [3], but the problem of the formal description of the Armenian language remained unsolved.

Another attempt was made in 2013-2014: the principles of the formal description of the Armenian derivative forms were developed [5]. Unfortunately, the program remained incomplete**.

Formal description of the Armenian vocabulary will enable to reveal regular structures, as well as variations, deflections, irregularities with their automatic analysis and the possibility of derivation [2]. It has an informative value, it can clarify and explain many unsolved problems in Armenian

*This project is supported by the Armenian Communities Department of the Galouste Gulbenkian Foundation.

**One of the performers of the above-mentioned project is the author of the following article. Therefore, implementing this work, the author used the previous valuable experience.

linguistics related to word formation. It will also provide much material for the further researches of Armenian grammar.

Material and methods. To do a formal description of the Armenian vocabulary we offer to solve *the following problems*:

1. studying the principles and methods of the formal description of the language used in modern linguistics to elaborate the general principles of the formal description of the Armenian word formation,
2. to determine the approximate volume of the vocabulary to be included in the database,
3. to select and elaborate appropriate word forms from dictionaries,
4. to do a formal description of the chosen word sequences from word formation point of view. Accordingly, simple, compound, derived and derived-compound words will be selected,
5. making word formation analysis of the abovementioned words, preparing the relevant linguistic material to develop the principles of the corresponding electronic database creating the database structure/template,
6. using all the possibilities of web technologies an online database will be created that will, in addition to a large database including thousands of words, allow the users to search simple, compound, derived and derived-compound words as well as stems, affixes and word endings. It will allow doing data comparison, observation, make comments and download linguistic data in different formats.
7. as an additional feature, stemming and lemmatization online tool [4] to will be available on the site.

The work should be done in *three stages*:

In the *first phase*, it is intended to select and elaborate the vocabulary to be included in the database, to do word formation description of the words by the following steps: to present the formal description of a word (main and secondary morphemes), to analyse word formation with consecutive steps, to present duplicates and their interpretations, to distinguish stems, to do word formation analysis of the stems, to present affixes and word endings.

In the *second phase*, the structure of the database will be created and finalized, the linguistic data will be entered. Only after all these steps when the linguistic data is complete, the website structure, its template, database structure, major functions of the site will be developed.

In the *third phase* it is planned to correct all possible mistakes and shortcomings of the website, to do scientific correction and editing of the linguistic material.

Research results. As a final result, we will have:

1. a huge electronic database available on its own official website that will allow users to search any word and get its word formation characteristic,
2. constructive analysis and formal description of thousands of words of the Armenian vocabulary, especially newly-created and widely used words,
3. a formal description of more than one hundred words. Later, in addition to a lexical database, many computer programs can be created, which will significantly contribute to computer processing of the Armenian language and vocabulary,
4. scientific articles, reports, and studies devoted to the word formation and formal description of the Armenian language.

The following electronic database will be unique in its kind with the volume of included vocabulary and will completely solve problems related to word formation of the Armenian language. An online available huge lexical database can serve a wide range of people, especially it can be useful to:

1. linguists, scientific researchers, scholars or students of philology faculties who want:
 - to study the Armenian word formation from synchronic and diachronic perspectives, to discover the basic patterns of word formation of new words,
 - analysing thousands of examples to find out the principles and ways of word formation patterns in the Armenian language,
 - to have a complete list of all simple, compound, derived and derived-compound words,
2. anyone, interested in the Armenian language with not enough linguistic knowledge, can search and find the word formation description of any word typed on the search bar of the website,
3. schoolchildren, applicants, teachers who do not know word formation well, but want to clarify, check any word form finding its stems, roots, affixes, morphemes and so on,

4. the database will also considerably support foreign Armenologists who are interested in the Armenian language as they do typological research.

The grammar system of the Armenian language is distinguished by the abundance of grammatical meanings and forms of expression, so in the first stage, it is planned only to do a formal description of the Armenian word formation. Only after the successful implementation of the current phase, it is planned to move to the next level of the language study. As a final result, we will have the full formal description of the Armenian grammar (morphology and syntax).

In addition to the electronic database created as a result of the formal description, it is also intended to create web and computer programs that will later serve to the computational processing of the language. Particularly those programs can be used to computationally process texts in linguistic corpora making analysis and distinguishing main and secondary forms of words.

Conclusions. Summing up, we can state that the following work, the formal description of the word formation of the Armenian language and the creation of the corresponding electronic database has a great perspective as it is the first phase of the complex program of the formal description of the Armenian language.

The number of users of the database will not be limited, as it will successfully serve not only linguists, students, schoolchildren living in Armenia and out of its borders, but also anyone who is interested in the vocabulary of the Armenian language and its word formation. The database may be useful to everyone with who has more or less knowledge of Armenian language, regardless of sex, age, educational level, place of residence.

The current database will greatly contribute to the full description of the Armenian language. Completing word formation formal description, it will be possible to move to the next level of the language study, first to the morphology, and then to the syntax.

Formal description of word formation of the Armenian language and creation of the corresponding electronic database will greatly contribute to the use of computer technology in linguistic research. This huge linguistic data will serve to various computer programs and highly contribute to the use of the Armenian language as a web language.

REFERENCES

1. Asilbekyan Sh. H., Formalized description of modern Armenian word order within UNL (2012); Grishkyan Y. H., Problems of description of meaning in machine translation theory (2013).
2. Asilbekyan, Sh.; Sargsyan, M., The Principles and Issues of the Creation of Formal Description of Armenian Language Vocabulary and Creation of the Corresponding Electronic Database. "Djahukian Readings", 2014, 6-11, Retrieved from http://language.sci.am/sites/default/files/book/jahowkryanakan_enterowmner.pdf
3. <http://www.unl.am/>
4. Lovins, J. B., Development of a Stemming Algorithm. "Mechanical Translation and Computational Linguistics 11", 1968, 22—31, Retrieved from <https://www.cs.waikato.ac.nz/~eibe/stemmers/index.html>
5. This work was made possible by research grant from the Armenian National Science and Educational Fund (ANSEF) based in New York, USA. THE 2014 ANSEF AWARDS
6. Urutyun R., Formal description of verb phrases in Modern Armenian, Yerevan, 1984, 165 pp; Hovsepian L., Problems and Principles of formal description of Modern Armenian language, "Jahukian Readings", Yerevan, 2007, pp. 112-117.