

ON THE ELECTRONIC DATABASE OF THE ARMENIAN WORD-FORMATION

MERI SARGSYAN

National Academy of Sciences of the Republic of Armenia,
International Scientific-Educational Center Scientific Secretary,
Institute of Language after H. Acharyan, Senior Researcher, PhD in Philological Sciences,
Associate Professor
meri.sargsyan@isec.am

<https://doi.org/10.52853/25792903-2021.2-msoe>

Abstract

The formal description of the languages has become more and more practical; however, it should be noted that the full formal description of the Armenian language has not yet been done. However, the fact that certain attempts have been made is undeniable.

The Electronic database of the Armenian word-formation (<https://formlang.am/>) is the first stage of the complex project in the full formal description of the Armenian language. In the current article, we want to present the advantages of the electronic database of the Armenian word-formation.

The electronic database containing the word-formation analysis of thousands of words can search for words and morphemes within them. It means that searching for any root or affix appears all the simple, compound, derived and derived-compound words made up of them. It enables us to reveal the regular structures and the variative forms, deflections, and irregularities with frequency data with their automatic analysis and the possibility of derivation.

The database gives a great opportunity to study the Armenian word-formation on synchronic and diachronic points, to discover the basic patterns formation new words, by the thousands of examples to find out the principles and ways of word-formation in the Armenian language, to have the full list of the distinguished simple, compound, derived and derived-compound Armenian words.

The current database has not only practical great value to involve the Armenian language in the domain of the modern informational technologies as the communicative mean, but also significant theoretical value to present the accurate description of the vocabulary structure. Thus, it will give an excellent perspective for solving the problems of theoretical linguistics and the practical -applied tasks. It can be significant for the further development of Armenology.

Keywords and phrases: formal description, word-formation analysis, Armenian language, electronic database, database query.

ՀԱՅԵՐԵՆԻ ԲԱՌԱԿԱԶՄՈՒԹՅԱՆ ԷԼԵԿՏՐՈՆԱՅԻՆ ՇՏԵՄԱՐԱՆԻ ՄԱՍԻՆ

ՄԵՐԻ ՍԱՐԳՍՅԱՆ

ՀՀ Գիտությունների ազգային ակադեմիայի
Գիտակրթական միջազգային կենտրոն,
գիտական քարտուղար,

ՀՀ Գիտությունների ազգային ակադեմիայի լեզվի ինստիտուտի ավագ
գիտաշխատող,

բանասիրական գիտությունների թեկնածու, դոցենտ
meri.sargsyan@isec.am

Համառոտագիր

Լեզուների ձևային նկարագրությունն ավելի ու ավելի կիրառական է դարձել, այնուամենայնիվ, հարկ է նշել, որ հայոց լեզվի լիարժեք ձևային նկարագրություն դեռ չի կատարվել, չնայած որ որոշակի փորձեր արվել են, անհերքելի է:

Հայերենի բառակազմության էլեկտրոնային շտեմարանը (<https://formlang.am/>) հայոց լեզվի լիարժեք ձևային նկարագրության առաջին փուլն է միայն: Սույն հոդվածում ներկայացրել ենք հայերենի բառակազմության էլեկտրոնային շտեմարանի առավելությունները:

Հազարավոր բառերի բառակազմական վերլուծություն պարունակող էլեկտրոնային շտեմարանը հնարավորություն է տալիս որոնելու բառեր, ինչպես նաև դրանց ձևոյթները: Սա նշանակում է, որ որևէ արմատ կամ ածանց փնտրելիս հայտնվում են դրանցից կազմված բոլոր պարզ, բարդ, ածանցավոր և բարդաձանցավոր բառերը: Այն մեզ հնարավորություն է տալիս բացահայտելու ինչպես կանոնավոր կազմությունների, այնպես էլ բոլոր տարբերակային ձևերի, շեղումների ու անկանոնությունների ամբողջական պատկերը՝ հաճախականության տվյալներով և դրանց ավտոմատ վերլուծության ու սերման հնարավորություններով:

Տվյալների շտեմարանը հիանալի հնարավորություն է տալիս ուսումնասիրելու հայերենի բառակազմությունը համաժամանակյա և տարժամանակյա հայեցակետերով, հայտնաբերել նոր բառերի առաջացման հիմնական օրինաչափությունները, հազարավոր օրինակներով պարզել հայերեն բառերի

կազմության սկզբունքներն ու ուղիները, ունենալ պարզ, ածանցավոր, բարդ և բարդաձանցավոր բառերի ամբողջական տարբերակված ցանկը:

Սույն տվյալների շտեմարանը ոչ միայն գործնական մեծ արժեք ունի հայոց լեզուն ժամանակակից տեղեկատվական տեխնոլոգիաների տիրույթում որպես հաղորդակցման միջոց ներգրավելու համար, այլև ունի տեսական մեծ արժեք բառապաշարի կառուցվածքի ճշգրիտ նկարագրությունը ներկայացնելու համար: Այն մեծ հեռանկարներ կստեղծի ոչ միայն տեսական լեզվաբանության, այլև գործնական-կիրառական խնդիրներ լուծելու համար: Այն կարող է շատ օգտակար լինել հայագիտության հետագա զարգացման համար:

Բանալի բառեր և բառակապակցություններ. ձևային նկարագրություն, բառակազմական վերլուծություն, հայոց լեզու, էլեկտրոնային շտեմարան, տվյալների շտեմարանի հարցում:

ОБ ЭЛЕКТРОННОЙ БАЗЕ ДАННЫХ СЛОВООБРАЗОВАНИЯ АРМЯНСКОГО ЯЗЫКА

МЕРИ САРГСЯН

Национальная академия наук Республики Армения

Международный научно-образовательный центр

ученый секретарь

старший научный сотрудник Института языка Национальной академии наук

кандидат филологических наук, доцент

meri.sargsyan@isec.am

Аннотация

Формальное описание языков становится все более практичным, однако следует отметить, что полное формальное описание армянского языка еще не выполнено. Однако тот факт, что были предприняты определенные попытки, неоспорим.

Электронная база данных словообразования армянского языка (<https://formlang.am/>) является первым этапом комплексного проекта по полному формальному описанию армянского языка. В данной статье мы хотим представить преимущества электронной базы данных словообразования армянского языка.

Электронная база данных, содержащая анализ словообразования нескольких тысяч слов, имеет возможность поиска слов, а также морфем. Это означает, что при поиске любого корня или аффикса появляются все простые, составные, производные и производно-составные слова, состоящие из них. Это позволяет раскрыть полную картину как регулярных формирований, так и всех вариативных форм, отклонений и

ненормативных форм с частотными данными, с возможностью их автоматического анализа и синтеза.

База данных дает прекрасную возможность изучить словообразование армянского языка в синхронном и диахроническом аспекте, обнаружить основные закономерности образования новых слов на тысячах примеров, выявить принципы и способы словообразования в армянском языке, иметь полный отличительный список простых, составных, производных и производно-составных армянских слов.

Существующая база данных имеет не только большое практическое значение для вовлечения армянского языка в сферу современных информационных технологий в качестве коммуникативного средства, но и теоретическое значение для точного описания структуры словарного запаса. Это даст большую перспективу для решения проблем теоретической лингвистики, а также практических и прикладных задач, что может быть очень важно для дальнейшего развития арменоведения.

Ключевые слова и словосочетания: формальное описание, словообразовательный анализ, армянский язык, электронная база данных, запрос базы данных.

Introduction

The formal description of the languages has become more and more practical; however, it should be noted that the full formal description of the Armenian language has not been done yet. Nevertheless, the fact that certain attempts have been made is undeniable [1]. The Electronic database of the Armenian word-formation [2] is the first stage of the complex project in the full formal description of the Armenian language. In the current article, we want to present the advantages of the electronic database of the Armenian word-formation.

The goal of creating this kind of database is analyzing thousands of words of Modern Armenian vocabulary, formal description of their components, in the case of compound words to reveal the complexity of components. The practical application of theoretical work is the creation of an online electronic database. To make the system available to the general public, the database runs on a separate website at **formlang.am** [2].

The objectives of the database

To reach the above-mentioned aim, the following problems have been solved:

1. The principles and methods of the formal description of the language used in modern linguistics have been studied to elaborate the general principles of the formal description of the Armenian word-formation [3].

2. Some of the most controversial issues of the Armenian word-formation have been studied, the disparities in professional literature have been analyzed, new observations and solutions have been presented [4].

3. The vocabulary to be included in the database was selected and elaborated.

4. We have determined the appropriate methods of the formal description of the word-formation doing word-formation description of the words by the following steps: to present the formal description of a word (main and secondary morphemes), to analyze word-formation with consecutive steps, to give duplicates and their interpretations, to distinguish stems, to do word-formation analysis of the stems, to present affixes and word endings.

5. All the affixes of Modern Armenian vocabulary have been presented. The affixes, with connecting vowel A and without it, have been presented separately. Now our database includes about 370 affixes. It should be noted that at the moment, there is no scientific source in which all Armenian affixes are mentioned; in this sense, this database is unique too.

6. The formal coincidences of interchanges and grammatical morphemes in the structure of affixes have been presented.

7. The classification of each affix in accordance with its position has been presented.

8. Also are presented the possible sound interchange, and the sound interchanged variants of each affix.

9. The principles of the corresponding electronic database creating the database structure/template have been developed.

We have also created the construct/framework of the database revealing the methodology principles of database creation; the software allows using the database in line with predefined operation conditions allowing finding needed data through the search function to realize data comparison, observation. To do all these, we have created a model that enables all users of the database of formal vocabulary descriptions to get the needed information.

The advantages of the database queries

The database queries enable us to get the complete picture of word-formation, exceptions, irregularities. Two types of searches may be expanded: simple and advances. Advanced search allows to make the following queries:

1. Search by word
2. Search by the first letter of the word
3. Search words with x stem
4. Search words formed with x derivational steps
5. Search words with x affixes

6. Search words with x roots and x word groups
7. Search x types of affix, prefix, suffix, infix and verbal affixes
8. Search words with sound interchanged roots
9. Search words with pure roots
10. Search words with x affixes
11. Search words with x roots / x word groups

For example, when we did an advanced search for a word կազդուրեցուցիչ, the following image appeared on the screen:

Lemma	կազդուրեցուցիչ								
Stem	կազդուրեցուց								
Stem Variative Form									
Prefix									
Roots	<table border="1"> <thead> <tr> <th>Root</th> <th>Pure</th> <th>Root Variative Form</th> </tr> </thead> <tbody> <tr> <td>կազնուր</td> <td>Այր</td> <td>կազնուր</td> </tr> </tbody> </table>			Root	Pure	Root Variative Form	կազնուր	Այր	կազնուր
Root	Pure	Root Variative Form							
կազնուր	Այր	կազնուր							
Verb Affix	եց								
Infix									
Connecting Vowel A	No								
Interfix Y	No								
Suffix	ուց իչ								
Ordinal Steps	3								
Notes	գրք: Կատոնարան/կան բառի հիմք (կազդուր+եաց+ուց) +իչ վրուց, ուրու եաց+եց, ուց+ուց								

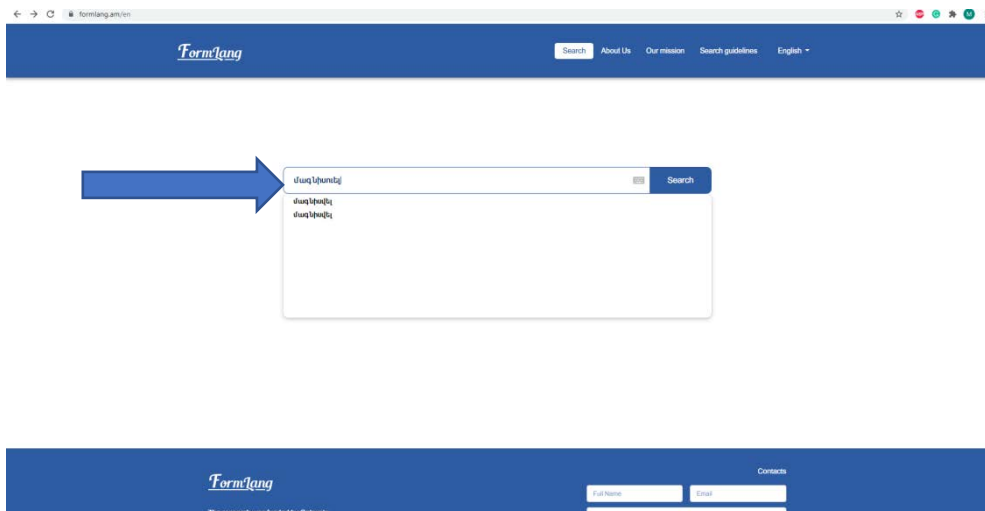
We updated the website that has new fields, new search capabilities with new and classic spellings.

The screenshot shows a web application interface with a search bar containing the letter 'մ'. Below the search bar, there are two checkboxes: 'Փնտրել վնասված բառերը' (checked) and 'Փնտրել չվնասված բառերը' (unchecked). To the right, there are two buttons: 'Անթացիկ Քառեր Word' and 'Անթացիկ Քառեր'. Below these elements is a table with the following data:

Մեկնաբանում	Բառը Պատմական Որոշագրությամբ	Բառի Քառաբանումը Վիճակ	Քառաբանման Դիմիկի Երկարություն	Նշումներ	Բառ. Բաղերի Ունիվ
մազազին	մազազին	մազազին		Փոխառություն	0
մազաբաթ	մազաբաթ	մազաբաթ			0
մազաբաթաբուր	մազաբաթաբուր	բուր			1
մազաբաթազար	մազաբաթազար	զար			1
մազաբաթազիլ	մազաբաթազիլ	զիլ			1
մազաբաթազրծ	մազաբաթազրծ	մազաբաթ			1
մազաբաթազրծուքրյան	մազաբաթազրծուքրյան	մազաբաթազրծ			2
մազաբաթազրյն	մազաբաթազրյն	զրյն			1
մազաբաթաթեթ	մազաբաթաթեթ	թեթ			1
մազաբաթաթուր	մազաբաթաթուր	թուր			1
մազաբաթաթար	մազաբաթաթար	թար			1
մազաբաթաթարմ	մազաբաթաթարմ	թարմ			1

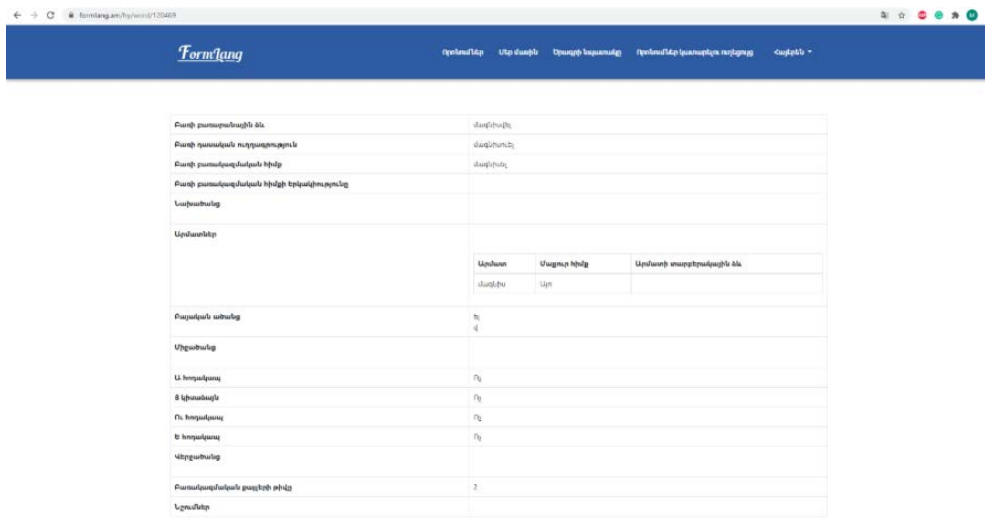
For example, we have added a new field in the database, searchable for the website users: words in Classical Armenian orthography. I.e. all the users can search words in both reformed and classical orthographies.

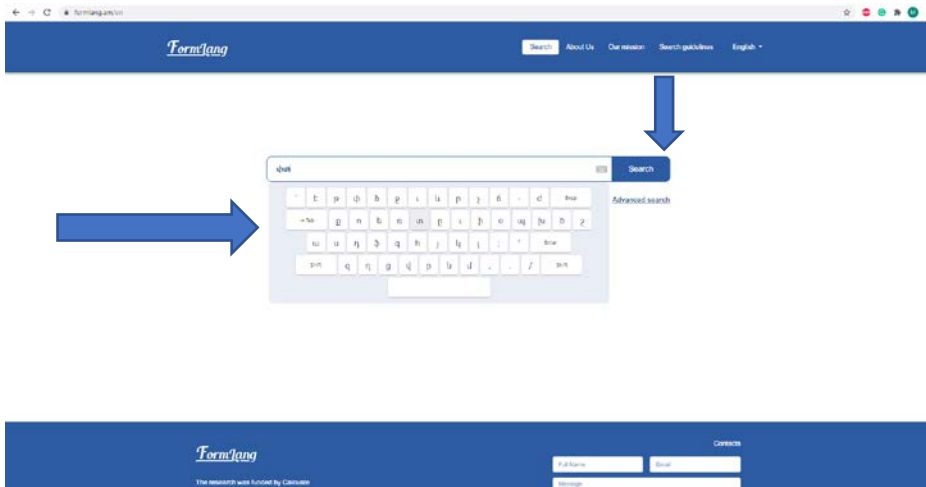
The database of the formal description of the Armenian language vocabulary will enable finding necessary data by the search system to realize data comparison and observation. The database queries will help to get the complete picture of word-formation, exceptions, irregularities. It will be possible to make the following queries (both classical and new spellings):



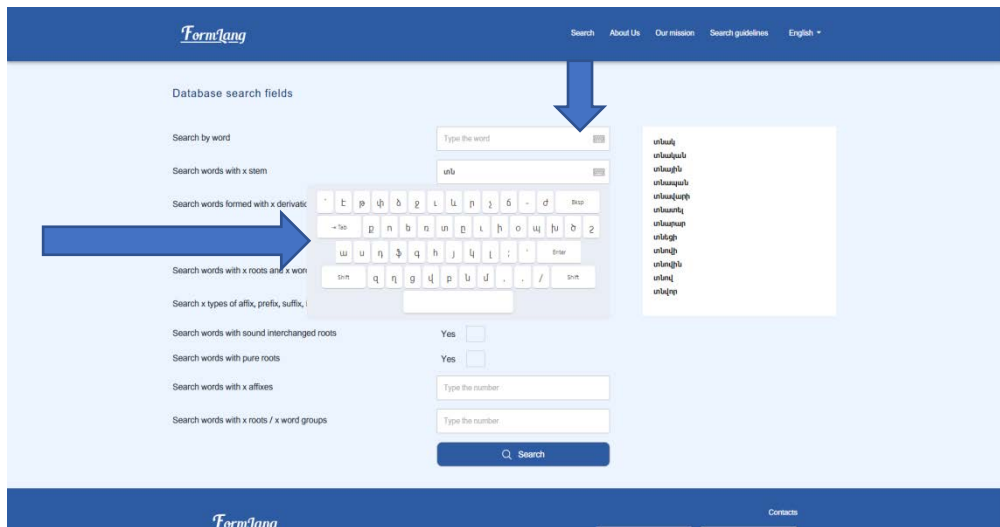
(Searching the word «մազկիսունէլ» in classical orthography)

We have developed and installed an Armenian keyboard in the system so that all users who have a technical problem writing Armenian can easily use the search engine.





The Armenian keyboard is available not only for regular search but also for advanced search. In addition, it is located in all search fields, which makes it quite easy to enter Armenian letters.



Conclusions

Being unique, this database and the search engine have theoretical and practical significance. The enormous linguistic material collected as a result of the formal description of the Armenian vocabulary allows those interested in studying the Armenian language's word-formation patterns from synchronic and diachronic perspectives.

The formal description of the Armenian vocabulary enables to reveal regular structures, as well as variations, deflections, irregularities with their automatic analysis and the possibility of derivation. Thus, the database gives a great opportunity to discover the basic patterns of the formation of new words, by the thousands of examples to find out the principles and ways of word formation in the Armenian language, to have the complete list of the distinguished simple, compound, derived and derived-compound Armenian words.

The current database has not only practical great value to involve the Armenian language in the domain of the modern informational technologies as the communicative mean, but also significant theoretical value to present the accurate description of the vocabulary structure. Thus, it will give an excellent perspective for solving the problems of theoretical linguistics and the practical -applied tasks. It can be essential for the further development of Armenology.

REFERENCES

1. Սարգսյան Մ. Ա., Հայերենի բառապաշարի ձևային նկարագրության կարևորության մասին, «Ա. Արրահամյանի ծննդյան 100-ամյակին նվիրված հանրապետական գիտաժողովի նյութերի ժողովածու», Ե., 2019, էջ 92-97:
2. Հայերենի բառակազմության էլեկտրոնային շտեմարան.
<https://formlang.am/>
3. **Sargsyan M. A.**, 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, January 2019, Warsaw, Poland, pp. 47-50.
4. Sargsyan M., Asilbekyan Sh., On some of the most controversial issues of the Armenian word-formation, Contemporary Issues of Historical Study of Armenian Language. Collection of Scientific Papers, 2019, Yerevan, pp. 177-185.

Հոդվածը ներկայացվել է տպագրության 015.06.2021,
ընդունվել է տպագրության 01.07.2021