



Для своего проекта я без раздумий выбрала ВАНК – Восточноармянский национальный корпус. У меня было огромное желание провести небольшое исследование армянского языка с точки зрения русского языка, так как я сама родилась в Ереване и знаю этот язык. Но мне действительно был интересен вопрос – насколько удобно проводить работы с корпусом армянского языка человеку, который плохо знает язык или вообще им не владеет.

Я была приятно удивлена, что в интернете очень много информации об этом малоизвестном корпусе. Есть ответы на все вопросы в различных статьях, сайтах. Но, честно говоря, я не увидела нужды в дополнительных ресурсах о ВАНК, так как все, что, возможно, вызовет вопросы, очень подробно объясняется на самом сайте.

### **Общая характеристика ресурса:**

- лингвистически репрезентативный корпус, насчитывающий около 110 млн. словоупотреблений
- мощный поисковый аппарат для осуществления сложных лексико-морфологических запросов
- удобный инструмент для самостоятельного изучения армянского языка, дающий для большинства словоформ лексико-морфологические разборы и английские переводные эквиваленты
- корпус, покрывающий различные периоды истории современного восточноармянского языка, начиная с 19-го века и до наших дней

# Восточноармянский и национальный корпус

корпус,

[HTTP://WWW.EANC.NET](http://www.eanc.net)

ГЕВОРКЯН ВИКТОРИЯ АРТУРОВНА, 1 ГРУППА,  
[VAGEVORKYAN@EDU.HSE.RU](mailto:VAGEVORKYAN@EDU.HSE.RU)

включающий как разножанровые письменные тексты, так и устные тексты разных типов

- аннотированный корпус, снабженный грамматической и библиографической разметкой
- корпус, находящийся в открытом доступе
- электронная библиотека, включающая более 100 классических произведений армянской литературы

### **Дизайн и общее впечатление.**

В целом, ВАНК очень напоминает всем нам известный НКРЯ, как в удобстве сайте, так и в поиске нужной информации. Мне сразу бросилось в глаза то, что ВАНК адаптирован для русскоязычных людей, а также есть версия на английском языке и, непосредственно, на армянском. Я воспользовалась версией на русском языке. Приятное отличие от НКРЯ – поиск и выдача отображены в одном окне, а подкорпус можно выбрать без первоначальной выдачи данных, а затем сохранения и возврата к поиску, как это устроено в НКРЯ. Сайт оформлен в очень простой и красивой гамме красного и оранжевого цвета, очень удобно, что основные советы и инструкции к поиску заключены вверху сайта, где и выбор языка сайта, что невозможно не заметить новичку. Основная информация о том, что же такое ВАНК, его характеристика, возможности и уровни разметки расположена слева, как и в НКРЯ. Каждая тема имеет свой заголовок и открывается в новом окне, чтобы не потерять предыдущую информацию. Также мне очень понравилось, что в левом меню указаны контакты, к кому можно обратиться с просьбой что-то улучшить или исправить в корпусе, указаны основные проекты и экспедиции, научные сотрудники Еревана и Москвы.

### **Ресурс глазами новичка.**

Сам восточноармянский корпус я нашла на сайте НКРЯ в разделе индоевропейских языков. При открытии ресурса можно легко заметить сам поиск нужной формы или леммы справа, не переходя в другие разделы. Также, о чем упоминалось выше, при проблемах с поиском новичок с легкостью разберется, перейдя по кнопке «КАК ИСКАТЬ», в которой не просто описано пошагово, как действовать, но и приведены примеры поиска со «скриншотами», что значительно всё упрощает. Очень понравилось то, что поиск можно задавать не только по подкорпусу, форме или лемме, но и по переводу слова. Например, если вы не владеете армянским языком, но вам нужно найти вхождения слова «любить» на армянском, то совсем бессмысленно искать это слово в русско-армянском словаре. Введя в поиске «love» на английском языке, ВАНК покажет вам все вхождения форм слова «sirel'», что означает «любить». ВАНК создан вовсе не для профессиональных лингвистов, в нем может разобраться любой «простой смертный». В ВАНК представлены все виды и пояснения разметок, также глоссарий, грамматические пометы, библиотека литературы и интернет-статьи на армянском языке. Честно сказать, работая с НКРЯ, возникало гораздо больше вопросов насчет поиска и самой выдачи. В ВАНК же все очевидно просто. Можно даже настроить выдачу так, как удобнее человеку. А все грамматические пометы отображаются при наводке курсора на слово в всплывающем маленьком окне.

## Список используемых в ВАНК лексико-морфологических помет

#	Помета ВАНК	Значение пометы	Армянский термин	Пример
		<b>Часть речи</b>	<b>Խոսքի մասեր</b>	
1	N	существительное	Գոյական	սեղան
2	A	прилагательное	Ածական	գեղեցիկ
3	V	глагол	Բայ	կարգալ
4	ADV	наречие	Մականջ	արագ
5	NUM	числительное	Թվական	երեք
6	PRON	местоимение	Դերանուն	ես
7	PREP	предлог	Կապ	սոսանց
8	POST	послелог	Կապ	մեջ
9	CONJ	союз	Շաղկապ	և
10	PART	частица	Ծղանակավորող բառեր	թերևս
11	INTJ	междометие	Զայնարկություն	վա՛յ
		<b>Части речи: дополнительные семантические признаки</b>		
12	S	независимое (субстантивное) местоимение	Անկախ դերանուն	ես
13	Dem	указательное местоимение	Ծուցական դերանուն	այդ
14	Intrg	вопросительное местоимение	Հարցական դերանուն	ինչ
15	Hum	личная референция (имя или местоимение)	Անձի անուն	մարդ
16	Anim	одушевленность (имя или местоимение)	Ծնչավոր	զառ
17	Inanim	неодушевленность (имя или местоимение)	Անշունչ	սեղան
18	Coll	коллективное имя	Հավաքական գոյական	խումբ
19	Topn	топоним	Տեղանուն	Հայաստան
20	Persn	имя	Անձնանուն	Արամ
21	Famn	фамилия	Ազգանուն	Պետրոսյան
22	Abbr	сокращение	Հապավում	ԱՊՀ
23	Card	порядковое числительное	Փասական թվական	երեք

Найдено: **123 013** вхождений, **7 029** документов

1. Վերք Հայաստանի Արովյան Խաչատուր 1841

Շատ փող սիրողի ջանին մանեթներ ա, որ կրակիցը հանում են, կպցնում...

2. Նոր Դար, 1

Մարինոն սուրբ կ

3. Աննա Կարենին

Ինչպե՞ս դու կար

4. Դեկամերոն

սիրել (V,tr)  
nmlz,ptcp,subj,sg,gen/dat  
love, have a passion/an affection for, like

սիրող (A)  
nmlz,sg,gen/dat  
amateur

Ես սիրել եմ քո աղջկան, սիրում եմ և պիտի սիրեմ ընդմիշտ, որովհետև նրան արժանի եմ հապա գործել եմ այնպիսի մի մեղք, որ միշտ էլ հատուկ է ջանել հասակին և որը վերացնելու հ

5. Ի սկզբանե էր բանն... Դավթյան Վահագն

Սիրո մեծությունը է չափվում նաև տաղանդի մեծությամբ, որովհետև սերը արարող է, և անա

## Помощь пользователю.

ВАНК имеет достаточное количество подсказок, инструкций и советов для нового пользователя, но все они собраны в ссылках и разделах на сайте, но подсказок при выполнении поиска и других функций, как, например, в НКРЯ – нет. То есть желательно сначала все прочитать, исследовать, разобраться, а затем приступать к своим задачам, потому что подсказки при поиске и выдаче данных отсутствуют. Но, по моему мнению, это не очень усложняет процесс, так как все предельно ясно описано в соответствующих разделах. К плюсам такой системы можно добавить, что поиск не нагроможден ненужной информацией для,

например, продвинутого пользователя. К тому же, все советы и инструкции в разделах расположены максимально удобно и просто, как и сама кнопка, где находятся все подсказки, которая называется «Как искать». Все предельно просто.

<b>КАК ПОЛЬЗОВАТЬСЯ КОРПУСОМ</b>
<b>ПОИСК СЛОВА</b>
<b>ПОИСК СОЧЕТАНИЙ СЛОВ</b>
<b>ПОДКОРПУС</b>
<b>ВИРТУАЛЬНАЯ КЛАВИАТУРА</b>
<b>ОТОБРАЖЕНИЕ РЕЗУЛЬТАТОВ</b>
<b>ПРИМЕРЫ ЗАПРОСОВ</b>
<b>ГЛОССАРИЙ</b>

## Как пользоваться корпусом

В этом разделе содержится введение в интерфейс ВАНК, которое можно использовать как руководство по первой очереди структуры пользовательского интерфейса и правил работы с ним; информация, необходимая

Раздел состоит из следующих подразделов:

[Поиск слова](#) – как искать отдельные слова.

[Поиск сочетаний слов](#) – как искать комбинации слов в одном контексте (т.е. контексты, одновременно содержа

[Подкорпус](#) – как ограничить область поиска той или иной категорией текстов (автором, временем создания, ж

[Виртуальная клавиатура](#) – как вводить армянские буквы, если на компьютере не установлена армянская клави

[Отображение результатов](#) – как просматривать найденные результаты, сортировать и расширять контексты, п

[Примеры запросов](#) – иллюстрации основных поисковых возможностей ВАНК с описанием последовательности

Названия элементов графического интерфейса ВАНК, которые используются в разделах 1 – 6, но понимание н


## **Сложные запросы.**

У рассматриваемого мной ресурса вполне обширная область функций, почти не отличающаяся от НКРЯ. Мне не очень понравилось, что при выдаче данных далеко не во всех статьях присутствует дата создания, а при многих исследованиях это очень важно. Лично мне, человеку, представляющему все-таки Россию, было бы удобнее работать с корпусом, если бы можно было искать леммы на русском языке, ведь пользователь не обязан знать английский. Также нет перевода выдачи данных. Не умеющему читать по-армянски пользователю будет не очень удобно работать с корпусом. Но можно настроить так, чтобы выдача отображалась латиницей.

Какие поиски выполняет ВАНК?


- Поиск словоформы
- Поиск лексемы
- Грамматический поиск
- Лексема + грамматический поиск
- Вхождения собственных имен
- Поиск сочетаний
- Поиск по подкорпусам
- Контекстные запросы
- Последовательности различных слов и так далее

Также в ВАНК нет оффлайн доступа, кроме скачивания выдачи. В корпусе очень удобно сравнивать употребления различных слов в разные эпохи, в разных точках мира. Это очень важно и нужно в армянском языке, так как диалектные и разговорные формы значительно отличаются от литературного языка. Например, я знаю пару диалектов армянского, прекрасно разговариваю в жизни, но прочитать и понять что-то из художественной литературы для меня очень проблематично. Поэтому это большой плюс ВАНКа. ВАНК умеет делать выдачи по нужному регистру букв, что важно при поиске имен, названий и т.д. На самом деле, корпус умеет искать абсолютно всё, что



**EANC**

Eastern Armenian National Corpus



ГЛАВНАЯ  
КАК ИСКАТЬ  
БИБЛИОТЕКА  
СТАТИСТИКА

---

Найдено: **3 160** вхождений, **1 289** документов

---

1. Նոթի Դաթը, **1901.09.20 #145**

Նրանց այս անհնայեղատ շարժումներն ազդում են ընտրի վրայ, որ աւելի է վահեցնում նրա բարեկամներին. սկսում են կանչել, նրա անունը տալ, որ գարջիմի, ասվակն ի դուր, նրանց բոլոր ջանքերին անցնում են ապարդյուն. կուում են ընկնելով հետ, որ նա այդպիսի բան արած:

[Расширить контекст](#)

---

2. Պատմվածքները Զոլյան Ստեփան

Ներքին չէր պատահաւում, եւ աշխրհ մորթի փական, մորթի բաց առատաւոյին հաւած, հնում էր. այտերը կարծես ուզում էին պաթել կարմրութիւնը... և այդ կարմրութիւնը վահեցնում էր Արաբին. նրան թվում էր, թէ երեսխանի ծանկնդով, փաթիլախով կարող է վանել այդ կարմրութիւնը, իսկ խոտեցնելով՝ սփառենգնել նրան ծանր թմրութիւնը:

[Расширить контекст](#)

---

3. Հակակաւան Ժամանակ, **2005.02.02**

Անաշաշտաւանության զգացողության աստիճանական ներշնչումը՝ սեփական երկրի ժողովրդին արարածինքնուր և հնազանդեցնելու, իսկ պտտենցիալ ներշնչողներին վահեցնելու՝ ամենալավ հանապարհն է:

[Расширить контекст](#)

---

4. Դեկամերոն Բուկաշտ Զիովանի 1960

Հապա փ գնանք տեսները այդ ինչ բորո է թեզ վահեցրի »:

[Расширить контекст](#)

---

5. Ոնիր և պատիժ Դոստոևսկի Ֆեոդոր 1980

Նա պարզ հիշում էր Լիզավետայի դեմքի արտահայտությունը այն պահին, երբ ինքը կացինը ձեռքին մտնում էր նրան, իսկ Լիզավետան նրանից ծա էր թաշվում դեպի պատը, ձեռքը դեպի առաջ պարզած, դեմքին բոլորովին ժանկական վախ, իսկ և իսկ ինչպես փոքրիկ երեխաները, երբ արանք հանկարծ սկսում են վախենալ որեւէ բանից, աշխարհ ու անհանգիստ կայուն են իրենց վահեցնող առարկային, եւ եւ թաշվում, թաթիկները դեպի առաջ պարզած, պատրաստվում են լաց լինելու:

[Расширить контекст](#)

---

6. Խաղաղ Դուրեյ, մաս 2 Շողիտով Միխայիլ 1977

— Վահեցնե՛մ դրանց, — վեր կենալով եղիպատարների ցողունների մաստառուտից, հարցրեց կոչնդը մյուսաներին:

[Расширить контекст](#)

---

7. Առաջադ, **2007.04.27**

[Расширить контекст](#)

Страница: Первая 1 2 3 4 5 6 7 8 9 10 ... Последняя
Версия для печати/MSWord | Сохранить в файл

#### 4. Պիեսանդ Օստրովսկի Ալեքսանդր 1987

ԿՈՒԿՈՒՇԿԻՆԱ — խոստովանանք լինի ասեմ, ես ամուրիներին նույնիսկ չեմ **սիրում** :

#### 5. Յո Նըթաս Սենկևիչ Հենդիկ 1983

Ափսոս, խանգարեց ինձ ազնիվ Վինիկիտաը, որ **սիրում** է նրան:

սիրել (V,tr)  
cvb,ipfv  
love, have a passion/an affection for, like

#### 6. Առանց հայրենիքի, Փյունիկ Ալաջաջյան Ստեփան

**Սիրում** են իրար, ծագում են իրար, ու գալիս է առիթը, և ահա մի բաժանում...

2. Запрос посложнее – Выберем подкорпус «Художественная литература» и зададим поиск по двум словам. Попробуем найти количество вхождений сочетания «գեղեցիկ երեխ» - «красивый ребенок». Будем искать в поиске леммы без признаков. Выдача показала все упоминания этого словосочетания из художественной литературы и все его формы.

**EANC**  
Eastern Armenian National Corpus

Найдено: **22** вхождений, **17** документов  
Размер подкорпуса: **37.11%** от общего объема БАНК

1. Դավիթ Բեկ Բաֆֆի 1882	<a href="#">Расширить контекст</a>
Նա այժմ մի առանձին երկուգածությունը նայում էր <b>գեղեցիկ երեխայի</b> երեսին, նայում էր դուրսով և չէր վստահանում կատարել այն, ինչ որ մտածում էր:	
2. Սամվել Բաֆֆի 1886	<a href="#">Расширить контекст</a>
<b>Գեղեցիկ երեխան</b> ուրախանալով ցույց էր տալիս, թե որքան անել էին «սիրունիկ» եղջուրները:	
3. Կովածաղիկ Պոռչյան Պեղն 1878	<a href="#">Расширить контекст</a>
— Ինչ <b>գեղեցիկ երեխոք</b> ունիս, Գևորգ խաչիս, — դարձավ դեպի տանուտերը թագավորը — ասոված քեզ քաշիս, քայց քանն այս է, որ փմյանց նման չեն. էս մի տղադ, — բռնելով Վարդի ձեռքը, — հակադրեց երկուսն էլ: Երբեք չես տեսնում, որ քա, էս մյուսը՝ շատ խելոք աչքեր ունի, իսկ աղջիկը միանգամայն Վրաստանի թագուհի կարող է դառնալ յուր գեղեցկությունով:	
4. Նըսա հանապալա՜ր, և Էնն	<a href="#">Расширить контекст</a>
— Բայց <b>գեղեցիկ երեխա</b>	
5. Ելլեք քույր Չելիով Անտոն 1973	<a href="#">Расширить контекст</a>

форма **лемма** перевод

գեղեցիկ 1

грамматика и части речи

Дополнительно ▾

Расстояние до следующего слова:  
от 1 до 1 (в словах)

форма **лемма** перевод

երեխա 2









грамматика и части речи

Дополнительно ▾

Расстояние: дополнительно ▾

Искать Очистить

3. Поиск с различными признаками (семантические отсутствуют в корпусе).  
 Попробую найти словосочетание, чтобы первое слово менялось, второе было собственным и не менялось. Подкорпус – художественная литература и устная речь. Год создания - до 2000г. Введем сочетание «աշքել Երեւանում» - жить в Ереване. Отметим, что «Ереван» должен быть с большой буквы и оставаться неизменным.

форма	лемма	перевод
шнур		 1
грамматика и части речи		
V,sg		
Дополнительно ▾		
Расстояние до следующего слова:		
от 1	до 1	(в словах)
форма	лемма	перевод
взрешены		 2
грамматика и части речи		
Пунктуация:		
слева	внутри	справа
		
Позиция в предложении:		
любая		
		
Регистр:		
с большой буквы		
		
Омонимичные разборы:		
разрешены		
		
Дополнительно ▲		
Расстояние: дополнительно ▾		
Искать	Очистить	

Таким образом, можно смело сказать, что ВАНК – достойный внимания корпус армянского языка. Конечно, в нем есть минусы, как и в любом другом ресурсе, но в целом, он достаточно адаптирован для не знающего армянский человека и очень прост в использовании. В нем приведен и состав текстов, и нужные разметки, и статистика слов и форм, и библиотека. ВАНК обновляется с каждым годом, поэтому надеюсь, что в поздних версиях появится долгожданный поиск по семантическим признакам, перевод выдач на русский язык и более расширенные ресурсы.



Статистика словоформ

На этой странице можно получить данные о частотности конкретной словоформы в корпусе в целом и для конкретных жанров/периодов в частности, полученные на основе частотного анализа текстов Восточноармянского национального корпуса. В приведенной ниже таблице даны три основные частотные характеристики трехсот самых частотных словоформ по корпусу в целом. Первое значение – это точное число словоупотреблений. Второе значение – это «число вхождений на миллион» (words per million), то есть среднее число вхождений данной словоформы на один миллион словоупотреблений корпуса. Наконец, третье значение – это ранг словоформы, стандартный показатель частотности, используемый в различных статистических приложениях и исследованиях и показывающий, насколько данный элемент частотен по сравнению с частотностью самого частотного элемента выборки ([подробнее](#)).

Чтобы просмотреть детализированную статистику частотности, щелкните по словоформе левой кнопкой мыши. Откроется окно, в котором дается точное число употреблений, число вхождений на миллион и ранг словоформы для корпуса в целом, а также все эти показатели для каждой из комбинаций декада / жанр. Статистика по текстам, дата создания которых неизвестна, приводится в предпоследней строчке. В последней строчке и последнем столбце приведены итоговые значения по всему жанру и декаде, соответственно.

Чтобы получить те же данные по словоформе, которая не входит в список трехсот наиболее частотных, введите ее в поле *Словоформа* и нажмите на кнопку *Показать статистику*.

Искать словоформу:   [Показать статистику](#)

Словоформы: 1-100				
#	Словоформа	Число вхождений	WPM	Ранг
1.	է	3,570,361	32374	1.0
2.	իմ	2,438,781	22114	1.5
3.	ըն	1,403,159	12723	2.5
4.	իմ	1,328,633	12047	2.7
5.	էն	1,187,500	10768	3.0
6.	ն	1,101,985	9992	3.2
7.	սի	852,799	7733	4.2
8.	ես	614,308	5570	5.8
9.	այդ	596,595	5410	6.0
10.	էլ	553,014	5014	6.5
11.	իր	501,075	4543	7.1
12.	ըն	494,885	4487	7.2
13.	համար	490,630	4449	7.3
14.	ես	485,047	4398	7.4
15.	իմ	406,493	3686	8.8
16.	էիմ	399,567	3623	8.9
17.	այն	392,437	3558	9.1
18.	այն	386,405	3504	9.2
19.	իմ	379 734	3439	9.4

Словоформы: 101-200				
#	Словоформа	Число вхождений	WPM	Ранг
101.	ըն	86,406	783	41.3
102.	էի	85,918	779	41.6
103.	ի	85,657	777	41.7
104.	այնիմ	85,183	772	41.9
105.	ընիմ	84,776	769	42.1
106.	ըն.р	84,590	767	42.2
107.	ըն.ես	84,023	762	42.5
108.	ը	81,287	737	43.9
109.	մի	81,067	735	44.0
110.	ի.ես	80,814	733	44.2
111.	հի.ես	80,372	729	44.4
112.	ի	79,585	722	44.9
113.	այն.ес	79,052	717	45.2
114.	ը.ес	79,023	717	45.2
115.	մ	78,548	712	45.5
116.	ը.այն.ес	77,259	701	46.2
117.	ը.ըն.հի.ес	76,974	698	46.4
118.	այ.ес	76,683	695	46.6
119.	и.ес	76.030	689	47.0

Словоформы: 201-300				
#	Словоформа	Число вхождений	WPM	Ранг
201.	ը.ес	47,395	430	75.3
202.	ը.и.ес	47,010	426	75.9
203.	ը.ес.р	46,691	423	76.5
204.	и.ес	46,004	417	77.6
205.	մ.ес	45,818	415	77.9
206.	ը.и.и.и.и.и.	45,568	413	78.4
207.	С	45,539	413	78.4
208.	и.и.и.и.и.и	45,467	412	78.5
209.	է.и	45,390	412	78.7
210.	10	45,360	411	78.7
211.	ի	45,330	411	78.8
212.	и.и.и.и.и.и.и	45,110	409	79.1
213.	и.и.и	45,085	409	79.2
214.	и.и.и.и	44,865	407	79.6
215.	и.и.и.и.и	44,617	405	80.0
216.	и.и.и	44,270	401	80.6
217.	и.и.и.и.и.и	44,042	399	81.1
218.	и.и.и.и.и	43,855	398	81.4
219.	и.и.и.и	43 771	396	81.7