

# Востоочноармянский национальный корпус



[www.eanc.net](http://www.eanc.net)

**Corpus Technologies**

# Армянские корпуса и электронные библиотеки:

## *Древнеармянский:*

[www.digilib.am](http://www.digilib.am)

[www.titus.uni-frankfurt.de](http://www.titus.uni-frankfurt.de) (корпус)

[www.sd-editions.com/LALT/index.html](http://www.sd-editions.com/LALT/index.html) (корпус)

## *Западно-армянский:*

[www.a.donabedian.free.fr/textes/Noojarmenien.pdf](http://www.a.donabedian.free.fr/textes/Noojarmenien.pdf)

[www.ermen.org](http://www.ermen.org) ...

## *Восточно-армянский:*

[www.people.cornell.edu](http://www.people.cornell.edu)

[www.armenianhouse.org](http://www.armenianhouse.org)

[www.arlis.am](http://www.arlis.am) ...

## *Армянская публицистика:*

[www.azg.am](http://www.azg.am)

[www.aravot.am](http://www.aravot.am)

[www.yerkir.am](http://www.yerkir.am) ...

## **Устная речь:**

Корпуса ВА устной речи отсутствуют

# Восточноармянский национальный корпус

- Около 110 млн. словоупотреблений
- Мощный поисковый аппарат
- Лексико-морфологические разборы и английские эквиваленты
- Тексты с XIX в. до наших дней
- Разножанровые письменные и устные тексты
- Грамматическая и библиографическая разметки
- Корпус, находящийся в открытом доступе
- Электронная библиотека (более 100 классических произведений армянской литературы)

# Этапы разработки

- ❑ **Март 2006 г.:** начало проекта
- ❑ **Июль 2007 г.:** первый релиз
- ❑ **Март 2008 г.:** второй релиз
- ❑ **Март 2009 г.:** третий релиз

# Состав ВАНК

Цель



максимально полный и диверсифицированный  
корпус текстов на литературном варианте  
восточноармянского языка

(с XIX века до наших дней)

⇒ *Не являясь эталоном языковой нормы, корпус призван отражать реальное использование языка*

## Состав ВАНК (на март 2009 г.)

Письменные тексты	словоупотр.	доля в ВАНК	документы
Художественная литература			
проза: романы	29 909 172	27,1 %	371
проза: рассказы	5 959 142	5,4 %	183
проза: драматургия	1 411 030	1,3 %	55
<b>итого прозы</b>	<b>37 279 344</b>	<b>33,8 %</b>	<b>609</b>
поэзия	3 648 160	3,3 %	227
Пресса	47 264 735	42,9 %	7858
Нехудожественные тексты			
научные тексты	13 875 930	12,6 %	113
эссе, мемуары, официальные и религиозные тексты	4 735 997	4,3 %	379
<b>Итого письменных текстов</b>	<b>106 804 166</b>	<b>96,8 %</b>	<b>9 186</b>

## Состав ВАНК (на март 2009 г.)

Устная речь	словоупотребления	доля в ВАНК	документы
Спонтанная устная речь	1 029 646	0,93%	208
Публичная устная речь	1 933 899	1,75%	543
Стимулированные нарративы	70 010	0,06%	22
+ Электронная коммуникация	442 399	0,40%	1
<b>Итого устный подкорпус</b>	<b>3 475 954</b>	<b>3,20%</b>	<b>774</b>
<b>Всего ВАНК</b>	<b>110 280 120</b>	<b>100%</b>	<b>9 960</b>

## Корпус письменных текстов

- ⇒ 106 млн. словоупотреблений
- ⇒ 510 авторов (с 1841 г. по 2009 г.)
- ⇒ 836 художественных произведений (в том числе 206 переводных)
- ⇒ 7858 выпусков периодических изданий
- ⇒ значительный объем научных и иных нехудожественных текстов



## Корпус устной речи

- ✓ Разные типы спонтанной речи
- ✓ Стимулированный дикурс («task-oriented»)
- ✓ Телевизионные программы
- ✓ Кинофильмы

+ Электронная коммуникация

---

## **Основные возможности поиска:**

- точный поиск словоформы
- поиск всех форм данной лексемы
- поиск по английскому переводу
- поиск по грамматическим и семантическим признакам
- лексема + грамматические и семантические признаки

## Дополнительные параметры поиска:

- Пунктуационные знаки
- Позиция в предложении (начало, середина, конец)
- Регистр
- Исключение грамматической (межлексемной и внутрилексемной) омонимии
- Логические функции и маски («и», «или», «\*» и т.д.)
- Выбор подкорпуса (по типу дискурса, жанру, автору, времени создания текста и т.д.)
- Подробная статистика словоформ

## Отображение результатов:

- ❑ Расширенный контекст
- ❑ Сортировка контекстов (по лексеме, словоформе, году создания, жанру и т.д.)
- ❑ Латинская транслитерация
- ❑ Глоссированная выдача
- ❑ Выдача KWIC (Key Words In Context)

# Пример транслитерации:

Найдено: **60 585** вхождений, **7 397** документов

1. Eresnamya kinə Balzak Ōnore də 1956

[Расширить контекст](#)

Et'e šnorhiv ayd tnayin davadrut'yan šat txmar **mardik** anc' en kenum orpes mec xelk'ov ōžtvac mardik, apa aydpisov havasarakš'vum ē xelok' mardkanc' k'anakə, oronk' himar en hamarvum:

2. Ařawōt, 03.28

[Расширить контекст](#)

Da nšanakum ē, or **mardik** patrast en patasxanatvut'yan irenc' bažinə verc'nel:

3. Ijewani koltntes, 08.11 #63

[Расширить контекст](#)

Hamašxarhayin čanač'um gtac amenatabrer profesianeri ter šat **mardik** en hangstac'el ays sanatoriayum:

# Псевдогlossированная выдача:

## 18. Ծաղիկներս մնացին հեռվում Վերոյան Բոգդան



խեղճ  
խեղճ (A)  
poor

**մարդիկ**  
մարդ (N,anim,hum)  
{pl nom}  
man

եք  
է (V,intr)  
{pres pl 2}  
be

գտել,  
գտնել (V,tr)  
{cvb pfv}  
find

ն'նց  
ննց (PRON,ADV,intrg)  
how

չէ:  
է (V,intr)  
{pres sg 3 neg}  
be

---

չէ (PART)  
no

# Выдача KWIC:

ած ժայռերի տակ կանգնած էին մի խումբ **մարդիկ** ՝ ուներին տոպրակներ, հագուստները տ

յնպես էլ սփյուռքում ավելի ու ավելի շատ **մարդիկ** ու կազմակերպություններ կտողորվեն այ

եվ այնտեղ **մարդիկ** լուր տվին դուքսին, /

Տգետ, անգրագետ **մարդիկ** արեվի խավարումը դիտում էին իբրև պ

Սովետական **մարդիկ** միշտ հասկացողութիւն մը եւ համակրանք

եզակի են **մարդիկ** , ում անունը կենդանության օրոք հավերժ **>>**

աշխատանքների սկիզբը, եղան այնպիսի **մարդիկ** , որոնք պնդում էին, թե Ասուանի բարձր

որ հենց սերն է՝ ստեղծվում է որ ապրեն **մարդիկ** :

ւմ ահա / Ա՛յն դանակի բթացումը, / Որով **մարդիկ** այգու նման կյանքն են էտում: /

**Մարդիկ** մեզ պրինցիպներից զուրկ են համարում:

# Кто является целевой аудиторией корпуса?

- ❑ Арменисты, филологи, типологи
- ❑ Студенты-филологи, школьники
- ❑ Преподаватели армянского языка
- ❑ Журналисты, редакторы ...



## Дальнейшие академические проекты ВАНК:

- ❑ Исследования на основе корпуса
- ❑ Глагольный словарь
- ❑ Грамматический словарь
- ❑ Диалектные корпуса ...

## Академическая поддержка проекта

Карен Акобян	Ереванский гос. лингвистический университет
Лилит Варданян	Университет Пизы
Люба Галстян	Институт бизнеса и политики (Москва)
Давид Гюрджинян	Ереванский гос. лингвистический университет
Анаид Донабебян	INALCO, Париж
Сурен Золян	Ереванский гос. лингвистический университет
Василиса Крылова	Санкт-Петербургский государственный университет
Левон Маркосян	Ереванский гос. лингвистический университет
Вера Подлесская	Российский гос. гуманитарный университет
Екатерина Рахилина	Институт русского языка РАН
Сусанна Тиоян	Ереванский гос. лингвистический университет
Роберт Урутян	Академия наук Армении
Фридрих Хлгатян	Ереванский гос. лингвистический университет
Нарине Экекян	Ереванский гос. лингвистический университет

# Востоочноармянский национальный корпус

[www.eanc.net](http://www.eanc.net)

Corpus Technologies



М. Даниэль, Д. Левонян, В. Плунгян, А. Поляков, С. Рубаков, В. Хуршудян