Восточноармянский национальный корпус

http://www.eanc.net

Геворкян Виктория Артуровна, 1 группа, vagevorkyan@edu.hse.ru

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 ресурcа:***

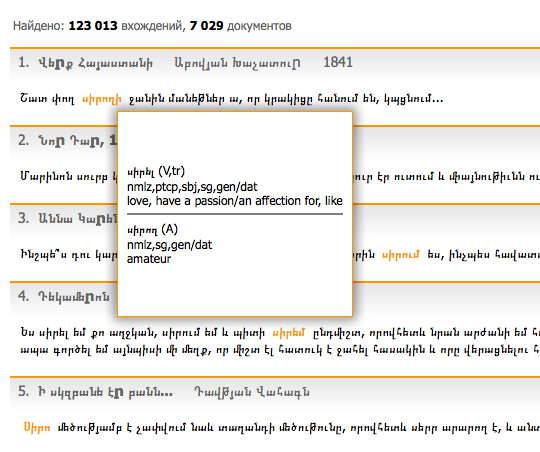
* *лингвистически репрезентативный корпус, насчитывающий около 110 млн. словоупотреблений*
* *мощный поисковый аппарат для осуществления сложных лексико-морфологических запросов*
* *удобный инструмент для самостоятельного изучения армянского языка, дающий для большинства словоформ лексико-морфологические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

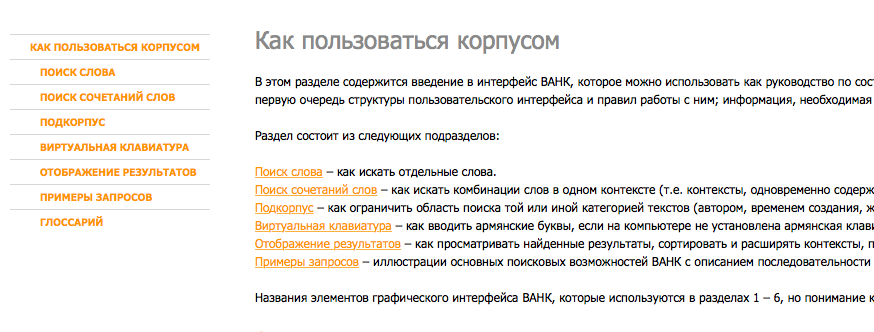
***Ресурс глазами новичка.***

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.



***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.



***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

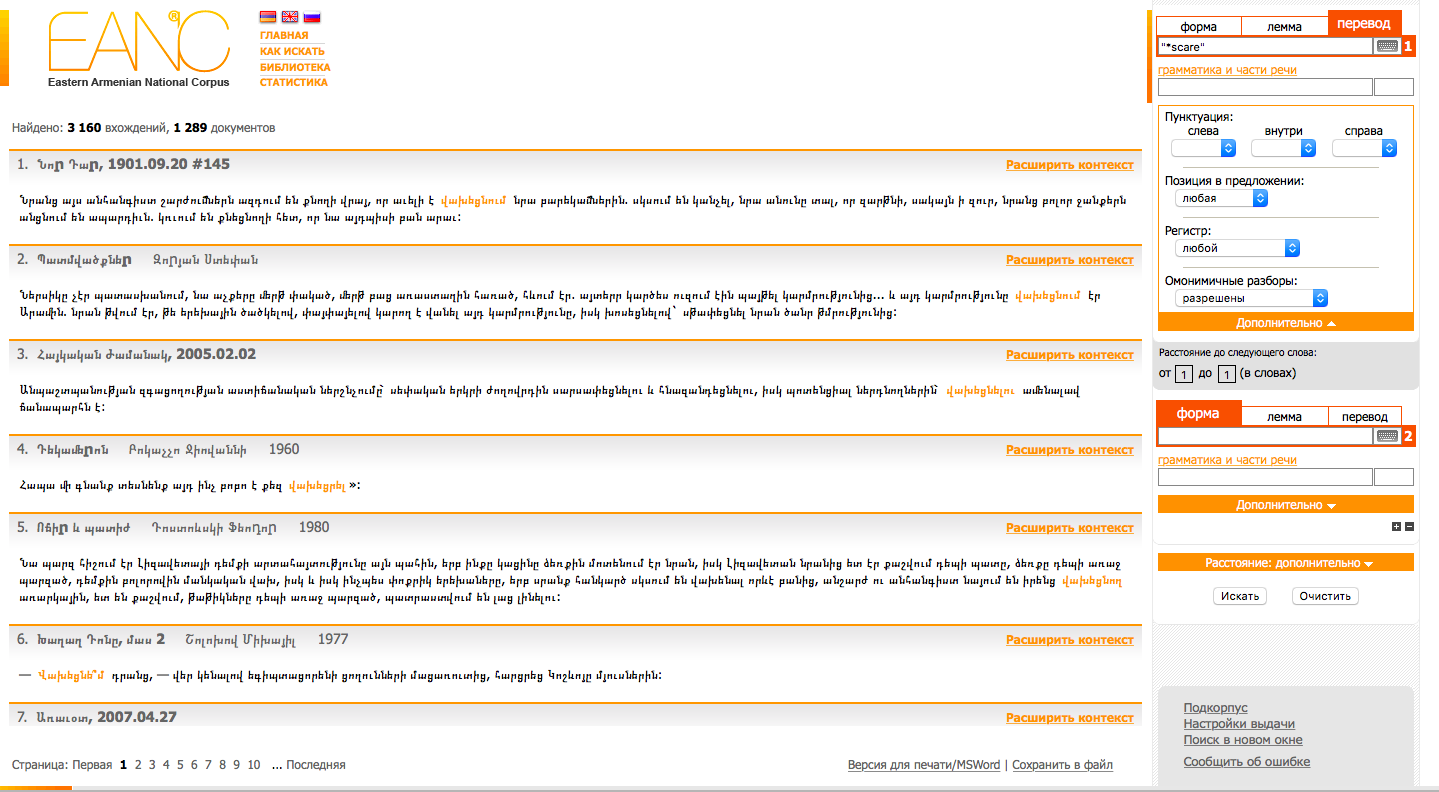
Также в ВАНК нет оффлайн доступа, кроме скачивания выдачи. В корпусе очень удобно сравнивать употребления различных слов в разные эпохи, в разных точках мира. Это очень важно и нужно в армянском языке, так как диалектные и разговорные формы значительно отличаются от литературного языка. Например, я знаю 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 понадобится юному лингвисту. Поэтому особые его «неумения» выделить достаточно сложно. Я лишь заметила, что очень мало соответствий при переводе с английского, если задавать фразовый глагол. В основном, он переводит, переводит корректно, но в очень частых случаях он выдает пустой лист. Без открытого глоссария сложно разобраться при выборе дополнительных и грамматических признаком заданного слова. В целом, функции ВАНК совпадают с НКРЯ.

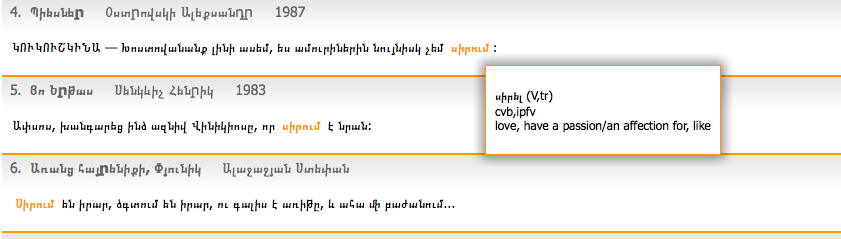
Нет никакого запроса по семантическим признакам, нет меню для обозначения нужных частей слова, как в НКРЯ. Это можно сделать только с помощью специальных обозначений, которые можно легко забыть.

***Пример запросов в ВАНК:***

Приведу пример нескольких запросов в данном корпусе.

1. Самый простой и обычный пример – откроем вкладку «перевод» и дадим запрос без каких-либо признаков слова «пугать» - «scare» и «любить» - «love» и обратим внимание на выдачу. Выдача при поиске в переводе будет состоять из всех найденных форм и однокоренных слов. Как вы видите на фотографиях ниже, если навести мышь на найденную форму или другое слово из выдачи – появятся его грамматические признаки и перевод на английский язык.



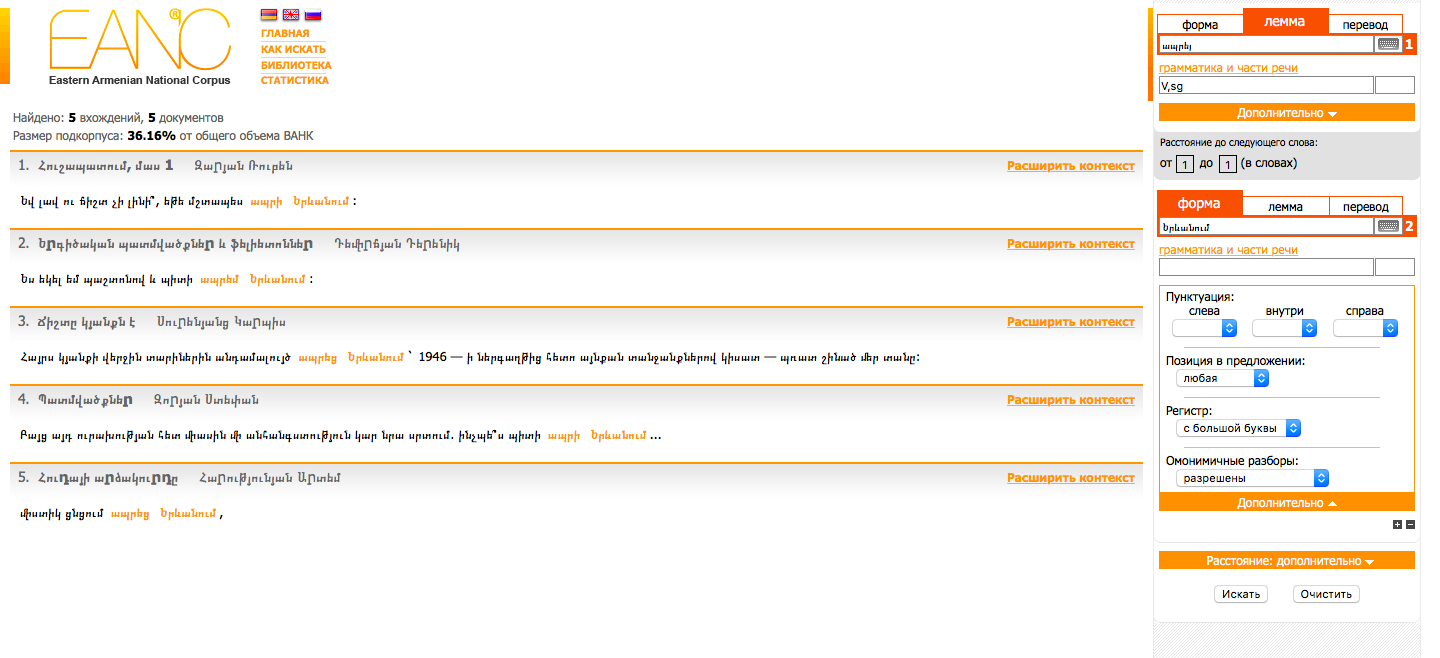


1. Запрос посложнее – Выберем подкорпус «Художественная литература» и зададим поиск по двум словам. Попробуем найти количество вхождений сочетания «գեղեցիկ երեխ» - «красивый ребенок». Будем искать в поиске леммы без признаков. Выдача показала все упоминания этого словосочетания из художественной литературы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



1. 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 «ապրել Երեւանում» - жить в Ереване. Отметим, что «Ереван» должен быть с большой буквы и оставаться неизменным.



Вхождений, конечно, мало, но они есть и даны действительно разные формы «жить» и одна форма «Ереване» и с большой буквы.

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.

