

## АДЫГЕЙСКИЙ КОРПУС И ПРЕПОДАВАНИЕ ЯЗЫКА

Ю. А. Ландер, НИУ «Высшая школа экономики», Москва, Россия, доцент, к. фил. н., адрес: 105066, г. Москва, Старая Басманная ул., д. 21/4, email: yulander@yandex.ru

И. Г. Багиорокова, Институт языкознания РАН / НИУ «Высшая школа экономики», Москва, Россия, научный сотрудник, доцент, к. фил. н., адрес: 125009, Москва, Большой Кисловский пер., д. 1, стр. 1, email: ibagiorkova@yandex.ru

Т. А. Архангельский, Университет Гамбурга, Гамбург, Германия, научный сотрудник, к. фил. н., адрес: Überseering 35, Postfach #29, 22297 Hamburg, Germany, email: timarkh@gmail.com

В работе описываются некоторые возможности Адыгейского корпуса – аннотированного электронного собрания текстов на адыгейском языке. В первую очередь нас интересует, как Адыгейский корпус может использоваться для преподавания и расширения представлений учащихся об адыгейском языке: в частности, речь идет как о составлении заданий, так и о сверке собственных представлений о языке с тем, как он реально используется. Более того, Адыгейский корпус может использоваться как источник для исследовательских проектов. В статье приводятся образцы заданий, которые могут быть предложены учащимся.

Ключевые слова: корпусная лингвистика, преподавание языка, Адыгейский корпус

## WEST CIRCASSIAN CORPUS AND LANGUAGE TEACHING

Yu. Lander, HSE University, Moscow, Russia, associate professor, Candidate of Sciences (PhD), 21/4 Staraya Basmannaya, 105066 Moscow, Russia, email: yulander@yandex.ru

I. Bagirokova, Institute of Linguistics RAS / HSE University, Moscow, Russia, research fellow, docent, Candidate of Sciences (PhD), B. Kislovskiy per. 1-1, 125009 Moscow, Russia, email: ibagirokova@yandex.ru

T. Arkhangelskiy, Universität Hamburg, Hamburg, Germany, research fellow, Candidate of Sciences (PhD), Überseering 35, Postfach #29, 22297 Hamburg, Germany, email: timarkh@gmail.com

This paper describes some aspects of the West Circassian Corpus, an electronic resource containing annotated texts in West Circassian (Adyghe). We focus on some ways in which the corpus can serve as an important instrument for teaching West Circassian. In particular, it is shown that the West Circassian Corpus can be used for compiling the exercises, for checking the actual use of the language as well as for organizing small-scale research projects devoted to the functioning of West Circassian. The paper contains some examples of the use of the West Circassian Corpus for teaching purposes.

Keywords: corpus linguistics, language teaching, West Circassian corpus

1. Электронные корпуса текстов – своего рода коллекции языкового материала, но коллекции особые. С одной стороны, их невозможно использовать как тексты для чтения, но зато в корпуса можно без риска для авторов и пользователей включать тексты, защищенные авторским правом. С другой стороны, корпуса обычно позволяют производить поиск внутри языкового

материала – как по конкретным лексемам и словоформам и их сочетаниям, так и по признакам словоформ (таким как «прилагательное», «единственное число», «прошедшее время» и т. д.). Это открывает перед тем, кто использует корпус (отнюдь не только лингвистом, но и литературоведом, фольклористом, историком, политологом и многими другими, в том числе и перед преподавателями языка), огромные возможности.

Совершенно естественно, что корпуса активно используются не только при изучении языка, но и при обучении языку. Для Национального корпуса русского языка ([ruscorpora.ru](http://ruscorpora.ru)) такое использование подробно описывается, например, в статьях [2; 3] и на специальном образовательном портале [studiorum-ruscorpora.ru](http://studiorum-ruscorpora.ru). Простейший пример такого использования – запрос в поисковой системе того или иного выражения для того, чтобы понять, используется ли это выражение (в интернете). Однако «продвинутые» корпуса позволяют и более интересное использование – своего рода исследование, позволяющее сделать занятие языком интересным (а иногда и неожиданным – в том числе для преподавателя). Другое существенное преимущество корпусов состоит в возможности ограничения текстов и исследовании фрагментов языка – газетной речи, диалектных текстов, художественной литературы (вплоть до работ конкретных авторов).

2. Адыгейский корпус (<http://adyghe.web-corpora.net/>), официально открытый осенью 2018 года, представляет собой особый ресурс, поскольку в отличие от многих других аналогов, он позволяет поиск по морфемам с учетом порядка морфем (см. о принципах организации корпуса [4; 5; 6]). Одновременно с этим в Адыгейском корпусе доступен и «классический» поиск – с помощью так называемых тэгов, отражающих грамматические признаки словоформы. Ниже мы кратко опишем лишь некоторые возможности корпуса.

Прежде, однако, следует подчеркнуть, что в силу того, что корпус должен содержать как можно большее количество материала (чтобы отражать язык в

достаточной мере), грамматический анализ, представленный в корпусе, производится автоматически – подобное количество текстов разобрать вручную попросту невозможно. В результате автоматического разбора помимо правильного анализа словоформ в корпусе представлен и неправильный. Хотя в каждой следующей версии корпуса число таких случаев уменьшается, оно остается большим, так что пользователь корпуса должен быть внимателен и не воспринимать результаты поиска как данность. Это следует воспринимать как вынужденную жертву, которая позволяет охватить большой языковой материал.

Общий интерфейс поисковой страницы Адыгейского корпуса показан на рис. 1.

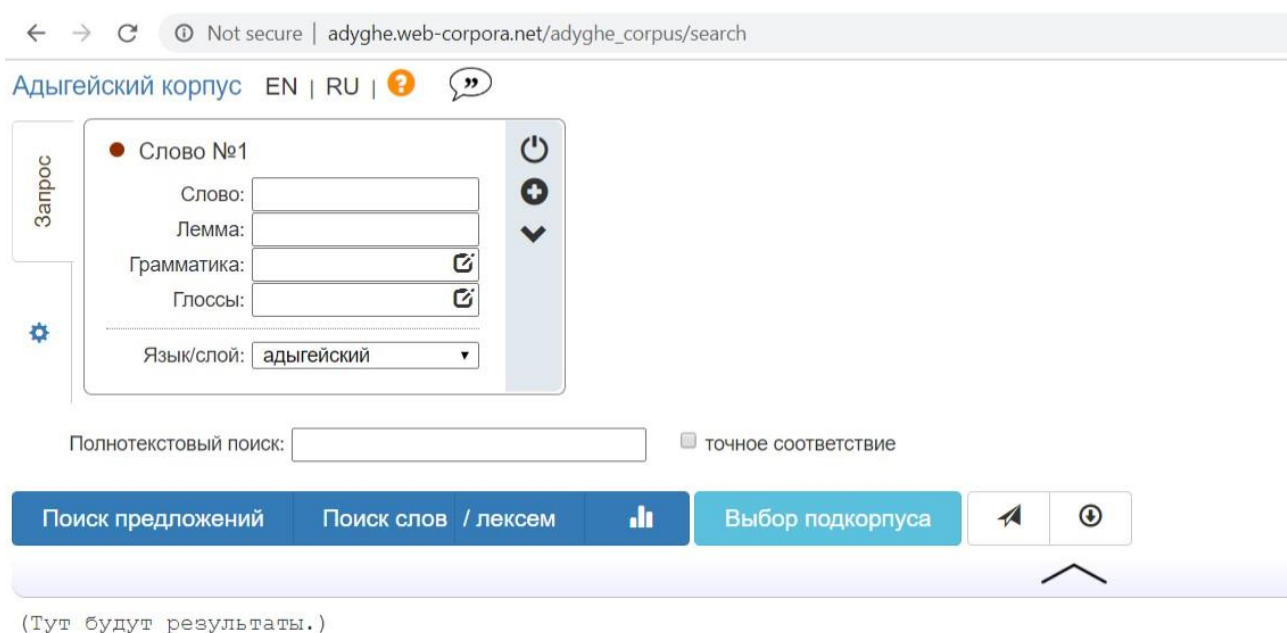


Рис. 1.

Запросы можно задавать в любом из полей, причем запросы в разных полях можно совмещать. Наиболее часто задействуются поля «Слово», «Грамматика» и «Глоссы».

Запрос «Слово» включает любую словоформу, причем любую ее часть можно заменить «звездочкой» (знак \*, который соответствует любой последовательности символов, в том числе нулевой), а любой отдельный символ (обязательно присутствующий) – вопросительным знаком. (Сразу оговоримся, что здесь и ниже мы указываем лишь один из способов запросов.) Например, запрос **сыкъ?\*бгъэ\*** соответствует слову, которое начинается на *сыкъ* (поскольку ему не предшествует \*) и содержит последовательность *бгъэ*, причем между этими последовательностями символов должен находиться хотя бы один знак (соответствующий ?\*). Среди результатов мы получаем предложения, которые содержат словоформы *сыкъызыфебгъэблэгъагъэм*, *сыкъытебгъэнагъ* и т. д. Символ «палочка» можно вводить разнообразными путями, в том числе и принятым в обиходе знаком единицы.

Запрос «Грамматика» включает грамматические признаки, список которых дан на заглавной странице (но также может быть вызван и со страницы поиска). Например, запрос **A, pst** в этом поле соответствует требованию, чтобы речь шла о прилагательном с показателем прошедшего времени (при этом, однако, во многих случаях выдаются словоформы, переводимые как прилагательные на русский язык – вроде *гъэнэфагъэкӀэ*).

Наконец, запрос «Глоссы» направлен на поиск последовательностей морфем, условные обозначения которых даны на заглавной странице. Морфемы разделяются дефисами. Любая последовательность морфем (в том числе пустая – отсутствие дополнительных морфем) заменяется «звездочкой», любая выраженная морфема заменяется знаком вопроса. Так, указание **neg-\*-pst-?-case** воспринимается как поиск любых словоформ, включающих показатель отрицания, следующий за ним на любом расстоянии показатель прошедшего времени (обязательно появляющийся после показателя отрицания, так что носитель адыгейского языка может понять, что отрицание в этом запросе

включает только префиксальное) и следующий за ним на расстоянии одной морфемы показатель падежа – вроде словоформы *имыхэбзагьэхэр*.

Искать в корпусе можно не только отдельные слова, но и словосочетания или даже предложения, которые просто содержат те или иные словоформы в любом порядке.

По состоянию на конец 2019 года Адыгейский корпус включает преимущественно газетные тексты (по большей части заимствованные с сайта газеты «Адыгэ макъ»); тем не менее в него включены также и многие произведения И. Машбаша, а также отдельные произведения некоторых других авторов. При обучении языку важно уметь ограничивать поиск. Это обеспечивается выбором подкорпуса, причем можно выбрать жанр, автора, время (иногда удобнее сперва ограничить жанр, а затем выбрать интересующие пользователя источники).

3. При обучении языку корпус используется, с одной стороны, преподавателем – при составлении упражнений и, с другой стороны, обучающимся – в первую очередь при выполнении проектов. Составление упражнений на те или иные явления достигается поиском примеров-решений в корпусе. Проекты, предлагаемые обучающимся, допускают значительный уровень фантазии, который может быть разным в зависимости от того, идет ли речь о школьнике или о студенте, и в зависимости от темы. Приведем простые примеры.

**Исследование использования понятий и лексем.** Рассмотрим контексты использования словоформ у И. Машбаша – в нашем случае словоформ лексемы *цIыфыгьэ*. Подкорпус ограничивается произведениями И. Машбаша, в строке «Слово» дается запрос **\*цIыфыгь\*** (подразумевается *цIыфыгьэ*, но допускается отсутствие последнего гласного, а также наличие дополнительных префиксов и

суффиксов). В ответ выдается более 170 примеров, на основании которых можно делать те или иные обобщения.

Так же исследуются и словосочетания, ограниченные какими-либо характеристиками. Например, нас может интересовать, с какими характеристиками в корпусе ассоциируются те или иные животные (строго говоря, речь идет об исследовании сочетаемости лексем с контекстами). Для этого сначала посмотрим, с какими прилагательными встречается наименование животного. Возьмем слово *мышьэ* 'медведь'. Форму **мышьэ** вводим в поле «слово» в разделе «Слово № 1» (заметим, что для некоторых слов в первом элементе атрибутивного комплекса может меняться гласный; например *баджэ* > *бэджэ*; см. обсуждение правила в [1]). Нажатием знака + открываются поля для дополнительного слова. Для этого слова мы задаем ограничение **A** (то есть «прилагательное») в строке «Грамматика». Через нажатие символа «добавить расстояние» (функция символа высвечивается при наведении на него курсора) накладываем ограничение, согласно которому это прилагательное должно непосредственно следовать за словом *мышьэ* – его расстояние от слова № 1 должно быть равно 1. Поиск в ответ на запрос, показанный на рис. 2, выдает все сочетания слова *мышьэ* с последующими прилагательными.

Рис. 2

Кроме того, можно посмотреть, кого сравнивают с медведями: в таком случае форму первого слова можно указать как **мышъ\*м** (чтобы учесть не только формы типа *мышъэм*, но и формы типа *мышъэхэм*), а второе слово специфицировать как **\*фэд\*** (чтобы включить формы типа *афэдэу*, *фэдэу* и т.д.). Надо сказать, что в обоих случаях на данный момент число результатов со словом *мышъэ* не очень велико – это обусловлено величиной Адыгейского корпуса. Поэтому возможно, такие задания лучше формулировать для списков лексем – хотя в будущем, хочется надеяться, количество текстов позволит проводить и более детальные изыскания.

**Исследования грамматики.** Даже не используя специальные грамматические пометы, на основе корпуса можно провести простейшие грамматические исследования, касающиеся сочетаний разных аффиксов, а также



аффиксов и корней. Например, с помощью простого запроса **\*шъхъапыры\*** мы получим набор словоформ, в котором встречается этот «сложный преверб» – последовательность морфем *шъхъа-пы-ры-*. При наличии должной сноровки можно проводить и более сложные исследования функционирования морфологии – находить разнообразные сочетания аффиксов, либо определяя их в полях «Грамматика» и/или «Глоссы», либо вводя ожидаемые последовательности в поле «Слово» - разумеется, используя звездочки и другие дополнительные символы. Точно так же можно исследовать функционирование слов – например, союзов вроде *ыкIи*. Задавая последовательности слов, ограниченные грамматическими признаками, можно изучать сложные грамматические конструкции.

Формулировка заданий и запросов, несомненно, требует от их авторов некоторой фантазии. Зато иногда в результате возникают весьма неожиданные находки. Так, запрос **neg-\*-neg** обнаруживает включающее два показателя отрицания слово *мыкIорэп* из поэзии И. Машбаша. В этой связи важно лишний раз подчеркнуть, что корпус содержит разнообразные примеры – отнюдь не только соответствующие строгой языковой норме, что имеет как плюсы (возможность проверить, что реально встречается), так и минусы (стилистически или диалектно специфические примеры, принимаемые не всеми носителями языка).

4. Итак, корпус представляет собой удобный инструмент для небольших исследовательских проектов, расширяющих знакомство учащихся с тем, как используется адыгейский язык. Помимо этого, как указывалось выше, в преподавании он может использоваться и для составления заданий, и для проверки того, встречаются ли те или иные выражения и конструкции.

Адыгейский корпус уже обладает достаточным размером и функционалом для многих сложных заданий, доступных тем, кто хотя бы на

каком-то уровне владеет адыгейским языком. Вместе с тем, для того, чтобы функционировать полноценно, Адыгейский корпус нуждается не только в постоянной доработке и исправлениях автоматического анализа словоформ – в процессе, естественном для любого корпуса, но также (и даже в большей степени) и в пополнении его новыми текстами, представляющими разнообразные жанры – в первую очередь, художественную прозу и фольклор. В этом авторы корпуса надеются на помощь тех, у кого есть доступ к таким текстам.

### Литература

1. Аркадьев П. А., Тестелец Я. Г. О трех чередованиях в адыгейском языке // Аспекты полисинтетизма: очерки по грамматике адыгейского языка. М.: РГГУ, 2009. С. 121—145.
2. Добрушина Н. Р. Как использовать Национальный корпус русского языка в образовании? // Национальный корпус русского языка: 2003—2005. М.: Индрик, 2005. С. 308—329.
3. Добрушина Н. Р. Корпусные методики обучения русскому языку // Национальный корпус русского языка: 2006—2008. Новые результаты и перспективы. СПб.: Нестор-История, 2009. С. 335—352.
4. Ландер Ю. А., Архангельский Т. А. Адыгейский корпус и орфографическое слово // Актуальные проблемы изучения кавказских языков. Материалы международной научно-практической конференции. Махачкала: Издательство ДГУ, 2017. С. 99—104.
5. Ландер Ю. А., Архангельский Т. А., Багиорова И. Г. Некоторые особенности адыгейского электронного корпуса // Кавказская филология: история и перспективы. К 90-летию Мухадина Абубекировича Кумахова. Нальчик: ИГИ КБНЦ РАН, 2019. С. 198—206.

6. Arkhangelskiy T. A., Lander Yu. A. Developing a polysynthetic language corpus: problems and solutions // Диалог 2016 / Компьютерная лингвистика и интеллектуальные технологии. 2016. № 15 (22). P. 38—47.