

Электронный корпус хакасского языка

Дизайн

В целом, сайт вызывает теплые и положительные впечатления, напоминая своим дизайном порталы начала 2000-х годов. Приятный серый цвет в сочетании со стандартным шрифтом Times New Roman совершенно не режет глаза и не мешает адекватному восприятию портала. Также все ссылки и кнопки совершенно стандартные, однако никаких проблем это не создает, так как они цветом (и в некоторых случаях расположением) выделяются из общей структуры сайта. Хочется отметить очень удачное расположение основных кнопок, которое сильно упрощает навигацию по сайту, делая его интуитивно понятным (а именно в строчку в самом верху). Из минусов можно отметить не очень сочетаемые цвета, применяемые при разметки корпуса (например, красный текст на бледно-зеленом фоне), что немного режет глаза, и не очень удачно сделана оптимизация для мобильных устройств (а именно при маленьком дисплее текст не очень удобно читать). При отсутствии каких-либо специальных подсказок сайт остается понятным, так как в каждом разделе есть пояснения для работы с ресурсом, которые достаточно обширны и информативны.

В целом портал выглядит достаточно приятно, хотя и стандартно.

Первый взгляд на ресурс

Электронный корпус хакасского языка доступен онлайн и его достаточно просто найти, ибо при запросах типа “корпус хакасского языка” в первых запросах поисковик выдает именно наш ресурс.

Хотя форма поиска достаточно проста, поиск присутствует только по грамматическим категориям. Причем найти что-либо на сайте достаточно просто (из-за сильно выделяющихся кнопок навигации) и образцы запросов просто не нужны (их и нет).

Электронный корпус хакасского языка

[Общее](#) |
 [Словарь](#) |
 [Грамматика](#) |
 [Тексты](#) |
 [Тексты \(в формате ЕАФ\)](#) |
 [Корпус](#) |
 [Диалектный подкорпус](#)
← ← Навигация по сайту

Параметры поиска ← Собственно поиск

Аффиксы (ручная разметка):

<input checked="" type="checkbox"/> 1SG	<input type="checkbox"/> ATTR	<input type="checkbox"/> DUR	<input type="checkbox"/> HAB	<input type="checkbox"/> NEG.CONV	<input type="checkbox"/> POSS3.ACC	<input type="checkbox"/> PROL	<input type="checkbox"/>	<input type="checkbox"/> СЮДА
<input type="checkbox"/> 2PL	<input type="checkbox"/> COMPT	<input type="checkbox"/> EMPH (ТОЖЕ)	<input type="checkbox"/> INF	<input type="checkbox"/> PAST	<input type="checkbox"/> POSS3.ACC	<input type="checkbox"/> PT	<input type="checkbox"/>	
<input type="checkbox"/> 2SG	<input type="checkbox"/> COND	<input type="checkbox"/> EVID	<input type="checkbox"/> INS	<input type="checkbox"/> PF	<input type="checkbox"/> POSS3.DAT	<input type="checkbox"/> REFL	<input type="checkbox"/>	
<input type="checkbox"/> ABL	<input type="checkbox"/> CONV1	<input type="checkbox"/> FORM	<input type="checkbox"/> INSTR	<input type="checkbox"/> PL	<input type="checkbox"/> POT.PT	<input type="checkbox"/> RPAST	<input type="checkbox"/>	
<input type="checkbox"/> ACC	<input type="checkbox"/> CONV1+ПРОДУЖЫВАТЬ=Р	<input type="checkbox"/> FORM.NEG	<input type="checkbox"/> INS.ДАЖЕ (EMPH)	<input type="checkbox"/> POSS1	<input type="checkbox"/> PRES	<input type="checkbox"/> ЖЕ(EMPH)	<input type="checkbox"/>	
<input type="checkbox"/> ALL	<input type="checkbox"/> CONV2	<input type="checkbox"/> FUT	<input type="checkbox"/> LOC	<input type="checkbox"/> POSS2	<input type="checkbox"/> PRES.ACT	<input type="checkbox"/> POSS3	<input type="checkbox"/>	
<input type="checkbox"/> APPOS	<input type="checkbox"/> DAT	<input type="checkbox"/> GEN	<input type="checkbox"/> NEG	<input type="checkbox"/> POSS3	<input type="checkbox"/> PREST	<input type="checkbox"/> CONV1	<input type="checkbox"/>	

Аффиксы (автоматическая разметка):

<input type="checkbox"/> 1SG.MIX	<input type="checkbox"/> 3POSS	<input type="checkbox"/> ACC.	<input type="checkbox"/> ATTR	<input type="checkbox"/> CONV.A	<input type="checkbox"/> DAT	<input type="checkbox"/> GEN	<input type="checkbox"/> INSTR	<input type="checkbox"/> NEG.CONV	<input type="checkbox"/> PL ₂	<input type="checkbox"/> RPAST
<input type="checkbox"/> 2POS.SG	<input type="checkbox"/> ABL	<input type="checkbox"/> APPOS	<input type="checkbox"/> COMPT	<input type="checkbox"/> CONV.NEG	<input type="checkbox"/> DUR	<input type="checkbox"/> HAB:	<input type="checkbox"/> LOC	<input type="checkbox"/> PAST:	<input type="checkbox"/> PRED.PL	
<input type="checkbox"/> 2SG.BR	<input type="checkbox"/> ACC	<input type="checkbox"/> ASS	<input type="checkbox"/> COND	<input type="checkbox"/> CONV.F	<input type="checkbox"/> FUT:	<input type="checkbox"/> IMP.1.INCL	<input type="checkbox"/> NEG	<input type="checkbox"/> PERF	<input type="checkbox"/> PRES	

Найдено фраз: 1.

Мини тайгада астых пар-ган кизе тоёйбін, хайдар парарга, хайди поларга пилбичем														
Мини я	тайгада тайга	астых булужная	пар-ган идти	кизе ПAST	кыз человек	ге DAT	тоёйбін похожий	исг	хайдар куда	пар-арга идти	хайди INF	пол-арга как	быть INF	пилбичем пнл-м бин че м знать исг FORM.NEG PRES исг
Я похож на человека, заблудившегося тайге, куда идти, что делать не знаю (ВИИ, ХТС, 116)														

← Выдана (виден пример не совсем удачного сочетания цветов)

Подводя итог, можно сказать, что данный портал сделан очень логично и понятно, и мы быстро находили необходимую нам информацию. Кроме того, ресурс сделан на русском языке, что сильно облегчило нам его использование.

На данный момент ресурс позволяет делать следующие вещи:

- Глоссировать хакасское слово/фразу

- Искать определенную грамматическую конструкцию

В данном случае мы искали конструкцию, в которой есть слово в прошедшем времени совершенного вида (из-за маленького количества текстов на большинство запросов корпус ничего не выдает)

Параметры поиска

Аффиксы (ручная разметка):

<input type="checkbox"/> 1sg	<input type="checkbox"/> COMIT	<input type="checkbox"/> EVID	<input type="checkbox"/> INSTR	<input type="checkbox"/> POSS1	<input type="checkbox"/> PRES.ACT	<input type="checkbox"/> CONV1
<input type="checkbox"/> 2PL	<input type="checkbox"/> COND	<input type="checkbox"/> FORM	<input type="checkbox"/> INS.DAT (EMPH)	<input type="checkbox"/> POSS2	<input type="checkbox"/> PRES.T	<input type="checkbox"/> CODA
<input type="checkbox"/> 2sg	<input type="checkbox"/> CONV1	<input type="checkbox"/> FORM.NEG	<input type="checkbox"/> LOC	<input type="checkbox"/> POSS3	<input type="checkbox"/> PROT.	
<input type="checkbox"/> ABL	<input type="checkbox"/> CONV1+ПРОДУМАТЬ+REFL	<input type="checkbox"/> FUT	<input type="checkbox"/> NEG	<input type="checkbox"/> POSS3.ACC	<input type="checkbox"/> PT	
<input type="checkbox"/> ACC	<input type="checkbox"/> CONV2	<input type="checkbox"/> GEN	<input type="checkbox"/> NEG.CONV	<input type="checkbox"/> POSS3.ACC	<input type="checkbox"/> REFL	
<input type="checkbox"/> ALL	<input type="checkbox"/> DAT	<input type="checkbox"/> HAB	<input type="checkbox"/> PAST	<input type="checkbox"/> POSS3.DAT	<input type="checkbox"/> RPAST	
<input type="checkbox"/> AP03	<input type="checkbox"/> DCL	<input type="checkbox"/> INF	<input type="checkbox"/> PF	<input type="checkbox"/> POT.PT	<input type="checkbox"/> X3(EMPH)	
<input type="checkbox"/> ATR	<input type="checkbox"/> EMPH (TOKE)	<input type="checkbox"/> INS	<input type="checkbox"/> PL	<input type="checkbox"/> PRES	<input type="checkbox"/> POSS3	

Аффиксы (автоматическая разметка):

<input type="checkbox"/> 1sg.AUX	<input type="checkbox"/> 3sg	<input type="checkbox"/> ACCS	<input type="checkbox"/> ATTR	<input type="checkbox"/> CONV.A	<input type="checkbox"/> DAT	<input type="checkbox"/> GEN	<input type="checkbox"/> INSTR	<input type="checkbox"/> NEG.CONV	<input type="checkbox"/> PL	<input type="checkbox"/> RPAST
<input type="checkbox"/> 3sg.SO	<input type="checkbox"/> ABL	<input type="checkbox"/> AP03	<input type="checkbox"/> COMIT	<input type="checkbox"/> CONV.NEG	<input type="checkbox"/> DCL	<input type="checkbox"/> HAB	<input type="checkbox"/> LOC	<input checked="" type="checkbox"/> PART	<input type="checkbox"/> PRES.P	
<input type="checkbox"/> 2sg.BV	<input type="checkbox"/> ACC	<input type="checkbox"/> ASS	<input type="checkbox"/> COND	<input type="checkbox"/> CONV.P	<input type="checkbox"/> FUT	<input type="checkbox"/> Imp.1.INCL	<input type="checkbox"/> NEG	<input checked="" type="checkbox"/> PERF	<input type="checkbox"/> PRES	

Поиск

Найдено фраз: 2.

– Че, чарир. Сия дее хаһыһыбыссан, паска полбас, – тии ала, Таргай стол хариңда одырыбыссан.

– Ну, хорошо. Если и ты управившись, неплохо, – говоря так, Таргай села около стола.

Сыдыдан тоопып, харах чызын чызып, Уклиңга сым на одырыбыссан.

Сыдыдан тоопып харах чызын чызып Уклиңга сым **на** одырыбыссан **одыр-мыбас хан** **склдет** **PERF PAST**

Закончив плач, вытерев слезы, Уклиңга села тихо.

Также в корпусе разрабатывается диалектный подкорпус, но пока он очень маленький (всего 1 аудиозапись). Вообще стоит отметить, что на данный момент корпус находится в стадии активной разработки, поэтому в нем всего порядка 150 тыс. словоупотреблений, что, несомненно, очень мало (в НКРЯ, например, около 283 млн словоупотреблений).

Корпус доступен онлайн, без регистрации. В нем доступна возможность скачивания текстов, но только в формате EAF (в диалектном подкорпусе также можно скачать аудиофайл и файл в формате .doc). По сравнению с НКРЯ корпус работает довольно быстро, однако иногда довольно долго загружает отдельные элементы.

Для корпуса литературных текстов размечены орфография, морфемное деление, глоссирование и перевод. В диалектном подкорпусе, помимо этого, указана транскрипция и звучание слова (в формате EAF есть возможность прикреплять звуковые файлы к тексту). Таким образом, корпус позволяет найти употребление отдельных грамматических форм, но выборка пока что нерепрезентативна. Тем не менее, разработчики собираются в ближайшее время сильно расширить функционал корпуса, доведя его до уровня НКРЯ. Так, например, словарь хакасского языка, который пока что есть только в starling, но который разработчики собираются добавить в корпус уже следующим обновлением, имеет отдельно разметку по семантическим признакам и словообразовательную.

Лимит выдачи ограничен 25 результатами, отсутствует синтаксическая разметка, нет возможности искать конкретную лемму или словоформу, контексты слов, можно искать только совокупность признаков, а не дизъюнкцию, но разработчики обещают все это в скором времени исправить.

Проект подготовили:

Мирон Крылов mironkrylov@gmail.com

Кирилл Денисов denisovkirill57@yandex.ru

Дмитрий Батов batoff.dmitry@yandex.ru

Ресурс: <http://khakas.altaica.ru>

Статьи с информацией про ресурс:

http://rodyaz.ru/pdf/no.5_2016/Дыбо%20А.В.%20Шеймович%20А.В.%20Аппарат%20автоматического%20морфологического%20анализа%20для%20корпуса%20хакасского%20языка.pdf

<https://textualheritage.org/content/view/405/188/lang.english/>