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).



Каких-либо специализированных инструкций ни на сайте, ни на внешних порталах обнаружено не было. При это ресурс содержит в себе обширные описания грамматических показателей и их значений, что помогает быстро разобраться человеку, даже совершенно незнакомому с хакасским языком.

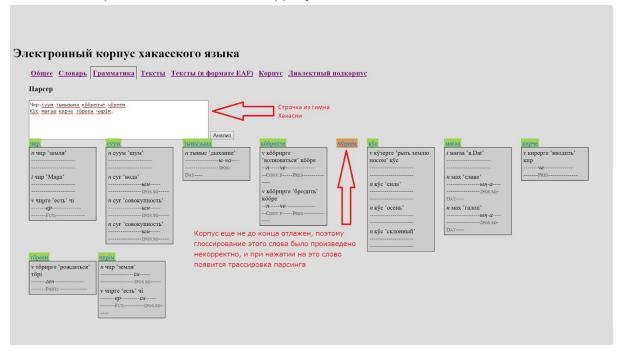
Из минусов можно отметить также то, что пояснения к поиску в корпусе находятся в другом разделе, а сам по себе поиск не всегда понятен интуитивно.

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.

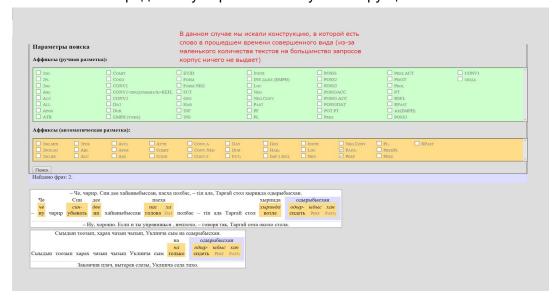
Продвинутый функционал

На данный момент ресурс позволяет делать следующие вещи:

- Глоссировать хакасское слово/фразу



- Искать определенную грамматическую конструкцию



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 <u>mironkrylov@gmail.com</u>
Кирилл Денисов <u>denisovkirill57@yandex.ru</u>
Дмитрий Батов <u>batoff.dmitry@yandex.ru</u>

Pecypc: http://khakas.altaica.ru

Статьи с информацией про ресурс:

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

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