

РОССИЙСКАЯ АКАДЕМИЯ НАУК
ИНСТИТУТ ЛИНГВИСТИЧЕСКИХ ИССЛЕДОВАНИЙ

**ЯЗЫКИ МАЛОЧИСЛЕННЫХ НАРОДОВ РОССИИ:
УСТНОЕ VS. ПИСЬМЕННОЕ**

Тезисы докладов
международной научной конференции
Санкт-Петербург, 4–6 декабря 2017 г.

Санкт-Петербург
2017

Языки малочисленных народов России: устное vs. письменное.
Тезисы докладов международной научной конференции,
Санкт-Петербург 4–6 декабря 2017 г. / СПб., 2017. — 68 с.

Редакционная коллегия: Е. В. Головки, С. А. Оскольская,
А. М. Певнов, А. А. Сюрюн, А. Ю. Урманчиева.

Издание осуществлено с оригинал-макета, подготовленного в ИЛИ РАН.
Оригинал-макет подготовила А. А. Сюрюн.

© Коллектив авторов, 2017
© ИЛИ РАН, 2017

Языки России в пространстве интернета: сайты, Википедия, социальная сеть

Б. В. Орехов

(НИУ Высшая школа экономики)

В 2016 году в НИУ ВШЭ завершился проект, целью которого был поиск и количественная оценка текстов на национальных языках народов России в Интернете. Все материалы (включая собранные корпуса) доступны в Интернете по адресу <http://web-corpora.net/wsgi3/minorlangs/>

Для поиска была разработана специальная методика, включавшая ручную работу с лингвистическими источниками (обследование грамматик и словарей) и создание программного кода для автоматических запросов к поисковой системе «Яндекс». Для языков, среди которых велся поиск, были найдены «слова-маркеры», позволяющие находить сайты, содержащие тексты на этих языках. Эти слова — единственный закрытый от свободного доступа результат проекта (во избежание вандализации списков).

К сожалению, тексты из Википедии для этой цели оказались непригодны: либо разделы на интересовавших нас языках были слишком малы, либо наполнены автоматически («ботозаливки»).

В структуру проекта входило также создание программных решений для скачивания текстов из социальной сети vk.com и для отделения текста на искомым языках от текстов на русском языке. Весь программный код открыт и доступен для повторного использования.

Проект предвзяли некоторые общие соображения, включавшие, например, такое: на родном языке пользователь социальной сети будет общаться, скорее, в специализированном сообществе, чем на личной странице. Кроме того: пользователь часто будет использовать не правильную орфографию, а тот режим, который мы вслед за А. А. Зализняком называем «бытовой системой письма», на практике означающий замену отсутствующих в гражданском русском алфавите графем на похожие, присутствующие на русской клавиатуре (ґ → г).

Во-первых, проведённое обследование позволило собрать некоторое количество текстов на языках народов России, тексты пригодны для включения в электронный корпус. В случае, если

текст извлекался из социальной сети vk.com, то для него также доступна богатая социолингвистическая разметка, включающая пол, дату рождения и город проживания говорящего. При этом данные анонимизированы и не сопоставлены с конкретным пользователем социальной сети.

Во-вторых, нам удалось составить примерную карту сетевого присутствия языков, включающую число найденных сайтов (доменных имён), сообществ в vk.com и статей в соответствующем разделе Википедии. Обнаружилось, что сетевые показатели витальности языка слабо коррелируют с числом носителей или другими количественными показателями, традиционно учитываемыми социолингвистами. При этом собственно коммуникация на национальном языке лишь в редких случаях оказывается «чистой»: в большом числе случаев даже в специализированных сообществах пишут по-русски больше, чем на собственно национальном языке.

Народная орфография негидальцев

А. М. Певнов

(Институт лингвистических исследований РАН)

Под народной орфографией имеются в виду индивидуальные, в разной степени упорядоченные опыты письменной фиксации родного языка теми людьми, у которых по разным причинам нет никаких преставлений об орфографии, разработанной в большей или меньшей степени на научной основе. Идеальные образцы народной орфографии мы находим у представителей тех народов, родные языки которых остались бесписьменными. К числу таких народов до недавнего времени относились негидальцы. В настоящее время негидальская орфография существует, однако применять её на практике фактически уже почти некому, так как негидальский язык находится на финальной стадии своего существования.

В докладе будут представлены образцы негидальской народной орфографии, полученные мной в 1985 г. в сёлах Ульчского района Хабаровского края от негидальцев, владевших языком своих предков. Материал был собран при помощи анкетирования, меня интересовала система негидальского вокализма. Письменные ответы моих респондентов нередко совпадали, что уже представляет