

Версии Википедии на приполярных языках

Тронд Тростеруд (*Trond Trosterud*), Университет Тромсё — Арктический университет Норвегии

Автореферат

Многоязычная энциклопедия Википедия, кажется, особенно хорошо закрепилась в приполярном регионе. Из приблизительно 70 приполярных языков, т. е. языков, являющихся коренными для данного региона, почти все имеют литературный язык, и у 45 из них есть собственная версия в Википедии или в инкубаторе Википедии. По сравнению с общим положением дел в мире, ситуация с Википедией и уровнем грамотности на крайнем Севере имеет уникальные особенности: согласно *ethnologue.com*, в мире существует 7117 живых языков, из которых более половины (3982) имеют литературный язык и лишь 620 имеют свою версию в основном пространстве или в инкубаторе Википедии. При более внимательном анализе версий Википедии на приполярных языках становится понятно, что эти цифры вводят в заблуждение. В настоящей статье исследуется, каким образом они вводят в заблуждение, а цель статьи состоит в том, чтобы лучше понять языковую грамотность коренных народов, ситуацию с приполярными языками в соответствующих странах и особенности создания небольших версий Википедии.

Введение

В настоящей статье будет проанализирован контент 45 версий основного пространства или инкубатора Википедии, написанных на языке приполярных меньшинств. Под *приполярным языком* понимается язык меньшинства, на котором говорит автохтонное население в районе к северу от 60-й параллели или близко к ней, за исключением языков, имеющих статус государственного.

Статья построена следующим образом: в разделе 2 представлены приполярные языковые сообщества, уровень грамотности в таких сообществах и версии Википедии на соответствующих языках. В разделе 3 анализируются различия между различными версиями на приполярном языке, причины, обуславливающие такие различия, а также роль, которую эти версии играют в возрождении языка. Раздел 4 резюмирует то, что мы теперь знаем об этих версиях и предпосылках, обусловивших успешность некоторых из таких версий, а в завершение приводится ряд рекомендаций касательно того, какие позитивные шаги можно предпринять в будущем.

Общие сведения

Приполярные языки

Если разделить приполярную область на три региона, то в североамериканском регионе присутствует 30 приполярных языков, в российском регионе — 31, а в странах Северной Европы — 10. Преобладающими языковыми семьями являются уральская (23 приполярных языка в странах Северной Европы и России), дене (19 приполярных языков в Северной Америке) и эскимосско-алеутская (10 языков во всех 3 регионах). Остальные 20 языков либо

являются изолятами, либо относятся к тюркским или тунгусо-маньчжурским языкам, за исключением германского фарерского.

Большинство приполярных языков имеет небольшое количество носителей; среднее количество носителей языков, рассматриваемых в данной статье, составляет 650 человек. 7 превалирующих языков (от 50 000 до 500 000 носителей) в рассматриваемой выборке распределены следующим образом (в порядке убывания по количеству носителей): якутский, тувинский, бурятский, коми, фарерский, коми-пермяцкий и гренландский. Каждый из них имеет собственную административную единицу, в которой язык носит статус официального. Такая единица может представлять собой либо автономную республику, либо область в России, либо остров с самоуправлением, как в случае с гренландским и фарерским языками.

В пределах своей автономной области три языка занимают доминирующее положение¹: фарерский (94%), тувинский (91%), гренландский (85%) и инуктитут (63%). Остальные языки являются языками меньшинств (якутский — 38%, бурятский — 27%, коми — 18%). Прочие языки имеют либо более слабый статус, либо вовсе не имеют официального статуса.

Типологически все приполярные языки имеют сложную морфологию. Языки на-дене и эскимосско-алеутские языки многосоставны, в них используется глагольное склонение для передачи грамматической и прагматической информации, которая в германских языках передается с помощью местоимений и наречий. В языках на-дене преобладают префиксные языки, а в эскимосско-алеутских — суффиксные. Приполярные языки, на которых говорят в России и странах Северной Европы, преимущественно являются суффиксными. Большинство этих языков имеют большие падежные системы (обычно от 6 до 10 падежей), но в языках коми и удмуртском имеется от 20 до 30 падежей. В языках меньшинств стран Северной Европы падежей меньше (менее 10), но они имеют более сложные системы чередования морфонологических основ при склонении. В целом, языки приполярных меньшинств заметно отличаются от языков большинства в соответствующих регионах. Наибольшей лингвистической удаленностью от языков большинства характеризуются приполярные языки в Северной Америке, а также гренландский и саамские языки в Норвегии и Швеции. В частности, русский и финский языки лингвистически менее удалены от языков меньшинств в своих странах.

Грамотность среди носителей приполярных языков

История вопроса грамотности на Севере восходит к трем различным инициативам по введению алфавита. Первая инициатива относится к миссионерской работе, проводившейся в период с 17 по 19 века включительно в связи с консолидацией государств. В результате такой деятельности в качестве основы для создания орфографии в странах Северной Европы был принят латинский алфавит, в России и на Аляске — кириллица, а в Северной Америке — силлабическое письмо и латиница. Вторая инициатива имела место только в России, где после Октябрьской революции была запущена крупномасштабная программа по плановому созданию литературных языков на основе латинского алфавита для всех языков меньшинств, включая северные, с идеей в конечном итоге заменить существовавшие ранее варианты орфографии на основе кириллицы. Наконец, в рамках процесса, начавшегося в 1960-х и 1970-х годах, первоначальные инициативы и последующее изменение или, по меньшей мере,

¹ Канада: [Statistics Canada 2016](#). Фарерские острова: [Statbank hagtova 2011](#), Гренландия: Frederiksen and Olsen 2017. Россия: [Всероссийская перепись населения 2010](#).

послабление политики ассимиляции привели к созданию новых видов орфографии для одних приполярных языков и к пересмотру старых вариантов орфографии для других.

Если внимательнее присмотреться к различным приполярным регионам, станет видно, что в Северной Америке на смену ранним миссионерским инициативам, которые реализовывались на языке коренных народов, начиная с 18-го века приходит ассимиляционная политика, усиливавшаяся по мере интеграции северных народов в современное общество в ходе 20-го века. В рамках политики примирения в 1999 году Канада отделила Нунавут от Северо-Западных территорий, предоставив собственный административный статус. В 2008 году Канада предоставила языку инуктитут статус официального языка и начала публиковать протоколы Законодательного собрания Нунавута на английском языке и языке инуктитут. Большинство приполярных языков в Канаде преподаются в резервациях коренных народов, но в основном они преподаются как иностранные языки. При этом обучение ведется на английском, и результаты изучения таких языков остаются неудовлетворительными (подробный анализ этой практики см. в Meek 2010). На Аляске ассимиляционная политика претерпела еще меньше изменений, чем в Канаде. В 1960-х годах Университет в Фэрбенксе начал создавать практическую орфографию для ряда языков коренного американского населения на Аляске, но это не привело к системным изменениям в школьной программе. Объявление 20 коренных языков Аляски «официальными языками наряду с английским» в 2014 году также не привело к каким-либо кардинальным сдвигам в политике².

В России создание орфографии в 1930-е годы сопровождалось административными изменениями, в рамках которых этнически «слепые» административные границы были заменены системой автономных республик, а в случае с более мелкими меньшинствами — областями, созданными вокруг всех меньшинств, достаточно крупных для образования административного органа. В первые десятилетия после Второй мировой войны в качестве языка обучения в начальных школах использовались языки, имеющие собственную автономную республику (коми, якутский, тувинский, бурятский, хакасский). В республиках издавались журналы и газеты на родном языке, литературные произведения также писались и публиковались на языке республик. Носителей других приполярных языков России обучали чтению и письму на родном языке, и на этих языках также выходили публикации. С введением новых законов о школьном образовании в 1958-1959 годах степень грамотности коренных народов снизилась, но по-прежнему оставалась на уровне, значительно превышающем уровень грамотности языковых меньшинств в других приполярных регионах³.

Фарерские острова на протяжении нескольких веков находились под влиянием датского языка, но в конце 19 века начали развивать собственную письменность. Британская оккупация во время Второй мировой войны ослабила влияние Дании, и после введения самоуправления в 1948 году фарерский язык постепенно стал доминирующим в том числе и в письменной форме (см. Knudsen 2010). В Гренландии гренландский язык использовался в

2 См. <http://www.akleg.gov/basis/Bill/Text/28?Hsid=HB0216Z>, новый пункт под названием «Adding the Inupiaq, Siberian Yupik, Central Alaskan Yup'ik, Alutiiq, Unangax, Dena'ina, Deg Xinag, Holikachuk, Koyukon, Upper Kuskokwim, Gwich'in, Tanana, Upper Tanana, Tanacross, Hän, Ahtna, Eyak, Tlingit, Haida, and Tsimshian languages as official languages of the state». В законе по-прежнему прямо говорится, что правительство штата и муниципалитет не обязаны печатать документ ни на каком другом языке, кроме английского.

3 Изложение советской политики одним из ее создателей — Исаев 1977. Для ознакомления с аналитическими материалами см., например, работы Kreindler 1985 и Trosterud 1995. В работе Bilinsky 1962 проводится анализ последствий введения законов о школьном образовании в 1958-1959 гг.

рамках всей системы школьного образования, включая подготовку педагогов, и даже в газетах, но в 1953 году эта вековая традиция местной грамотности была разрушена, когда Гренландия превратилась из датской колонии в округ (*amt*), являющийся частью Дании. Эти преобразования сопровождались масштабным «смещением» грамотности (переходом с гренландского на датский). Несмотря на то, что введение внутренней автономии в 1979 г. и самоуправления в 2008 г. сделало гренландский язык единственным официальным языком Гренландии, датский язык по-прежнему занимает устойчивое положение, особенно в качестве доминирующего языка во всей системе академического образования выше обязательного уровня. В сфере административной, правовой системы и средств массовой информации применяется концепция двуязычности — текст обычно составляется на датском языке и переводится на гренландский (Frederiksen and Olsen, op. cit.).

За последние четыре десятилетия страны Северной Европы постепенно внедрили новую языковую политику для своих языков меньшинств, в рамках которой такие языки получили официальный статус на региональном уровне или, в случае с саамскими языками, квенским и меянкиели, статус языка национальных меньшинств. Большая часть учебной программы в школах северных саамов преподается на северносаамском и в меньшей степени на колтта-саамском и инари-саамском языках. Все языки приполярных меньшинств в странах Северной Европы преподаются в качестве школьных предметов.⁴ Следует отметить исключительный случай с инари-саамским языком, когда в результате языкового перехода, наблюдавшегося в течение последнего поколения, сложилась ситуация, когда большинство носителей принадлежит к молодому поколению, и журналы, учебники, художественная литература и новости выходят на инари-саамском языке⁵.

Итак, для приполярных языков в целом ситуация складывается таким образом, что только четверть нынешних орфографий относятся к периоду до Второй мировой войны. Эти орфографии либо кириллические, либо основаны на орфографии 1937 года, либо представляют собой слоговое письмо. Возраст половины приполярных орфографий составляет менее 40 лет. Основная часть носителей большинства приполярных языков принадлежит к старшим возрастным когортам, поэтому типичная ситуация складывается таким образом, что орфографии приполярных языков младше, чем их собственные носители.

Что касается использования языков в социальных сетях, в Facebook можно найти как фарерский, гренландский, меянкиели, северносаамский, так и инари-саамский, а также в определенной степени колтта-саамский, южносаамский, луле-саамский, карельский, ливвиковское наречие и вепский язык — как в группах, посвященных соответствующим языкам, так и в личных публикациях. В Северной Америке ситуация выглядит иначе, поскольку во всех группах в Facebook, посвященных самым распространенным языкам (инуктитут, центрально-аляскинский юпик, дене сулин, инупиак), используется только английский язык. Слова на родном языке обычно берутся в кавычки (например, «что означает X»), но обсуждение всегда ведется на английском языке. Бурятский, коми, тувинский и хакасский языки используются в группах vkontakte⁶, посвященных использованию данных

4 Чтобы получить представление о политике Норвегии, см. План действий в отношении саамских языков (Министерство труда и социальной интеграции Норвегии, 2009 г.). Анализ ситуации в области саамских языков в Финляндии представлен в работе Pasanen 2016.

5 Для получения информации о процессе возрождения языка можно обратиться к часто упоминаемой работе Pasanen 2015. Углубленный анализ восстановления «среднего поколения» (поколения между старшими носителями языка и детьми) представлен в работе Olthuis et al. 2013.

6 vkontakte — российский аналог Facebook, см. <https://vk.com/>

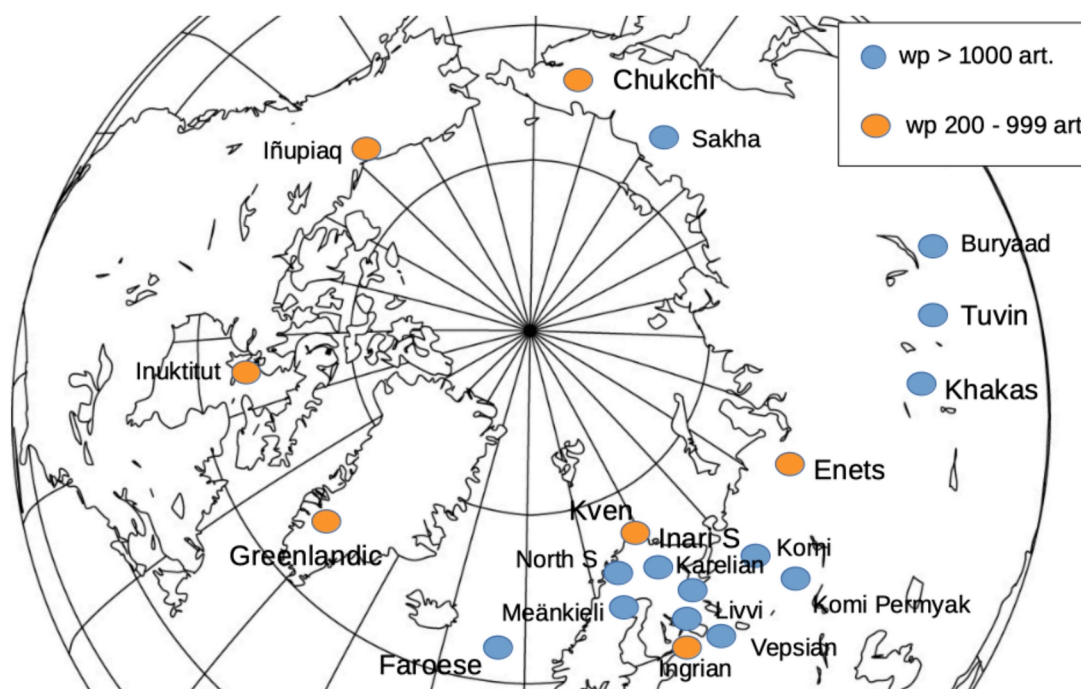
языков; якутский язык — в меньшей степени. Что касается других приполярных языков России, имеющих собственную версию Википедии, то складывается впечатление, что их языковые группы вkontakte ведутся на русском языке, либо отсутствуют вовсе.

Версии Википедии на приполярных языках

Теперь перейдем к конкретным версиям Википедии. Проведя краткий обзор, рассмотрим размер версий и их языковое сообщество, затем авторов и, наконец, состояние статей.

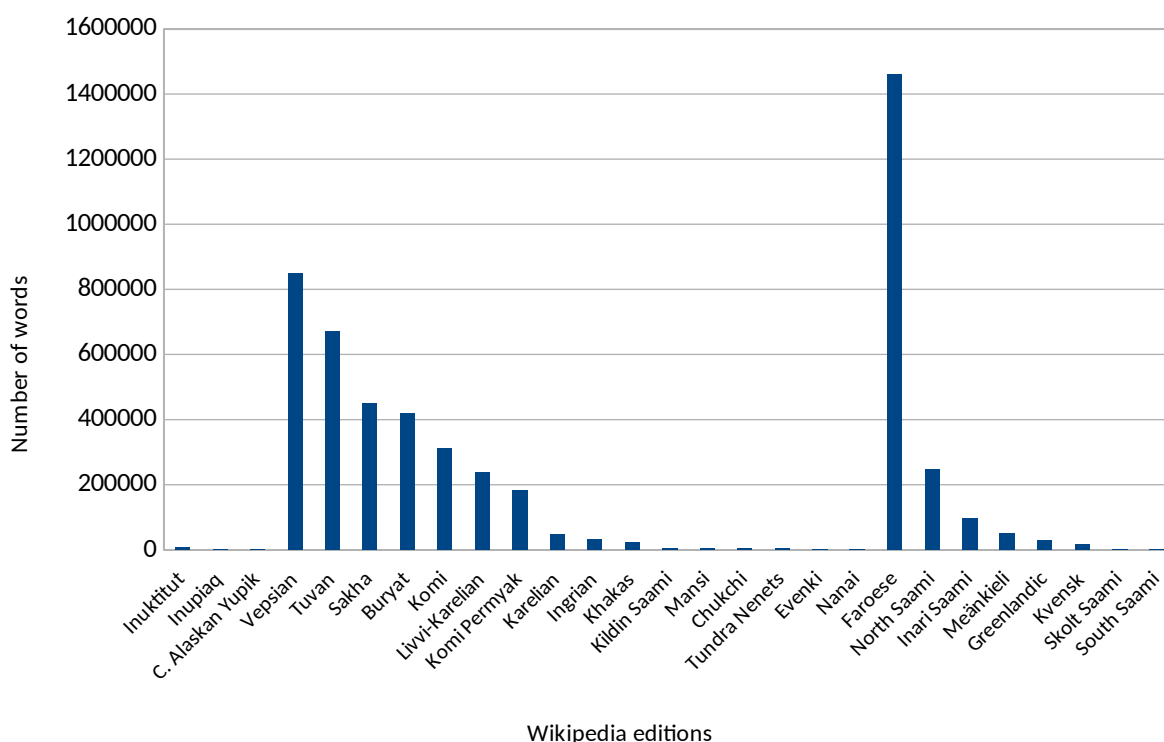
Версии Википедии

Карта на Рисунке 1 показывает примерное расположение версий Википедии, имеющих более 200 статей (отображены оранжевыми овалами), и версий, имеющих более 1000 статей (отображены синими овалами). Как видно из рисунка 1, версии Википедии, содержащие более 1000 слов, встречаются в основном в странах Северной Европы, но имеются и в России, будучи составленными на республиканских языках. Помимо языков, показанных на карте, существует 24 версии Википедии, содержащие менее 200 слов. Если идти от большего к меньшему, то распределение выглядит следующим образом: 6 относятся к Северной Америке (центрально-алаяскинский юпик, тлингитский язык, хэн, гвичин, алеутский и дене сулин), 15 — к России (кильдинский саамский язык, мансийский, кетский, тундреный ненецкий, лудиковский, тундреный юкагирский, нанайский, эвенкийский, сибирский юпик, хантыйский, долганский, удэгейский, ульчский, ороцкий и негидальский) и к странам Северной Европы (южносаамский, колтта-саамский и луле-саамский). Для языков этой группы характерно то, что они имеют мало носителей, а литературные традиции слабы или вообще отсутствуют. Версии Википедии на таких языках обычно создаются людьми, которые не являются их носителями и не входят в соответствующее языковое сообщество, некоторые статьи написаны на языке большинства, и в настоящее время ни в одной из них не наблюдается никакой активности.



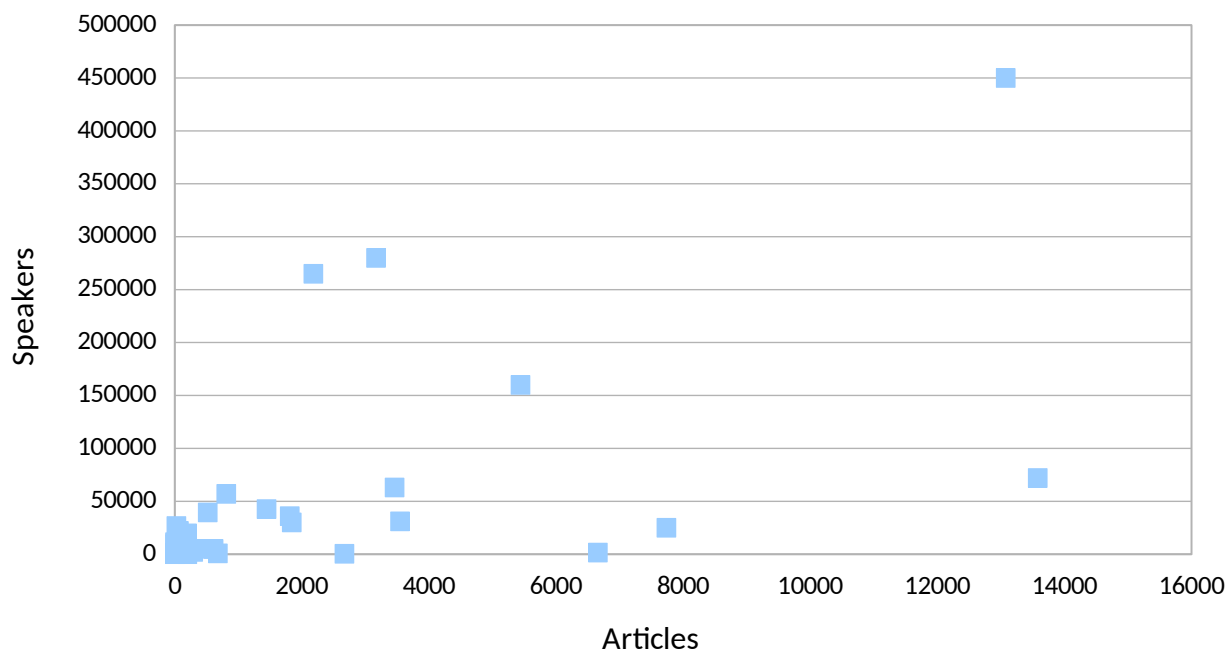
Размер Википедии

Прежде всего оценим размер версий Википедии, содержащих более 1000 слов. Они представлены на рис. 2 и упорядочены сначала по регионам (Северная Америка, Россия, страны Северной Европы), а затем по количеству слов. Как можно заметить, эти три региона имеют характерные особенности. Что касается Северной Америки (слева), то в трех крупнейших версиях Википедии едва ли насчитывается 100 слов. Версии Википедии в России (за некоторыми исключениями, к которым мы еще вернемся) можно разделить на две группы: те, у которых есть поддержка в лице собственной республики, и те, у кого ее нет. В странах Северной Европы особняком стоит фарерский язык. Северносаамский язык является наиболее используемым после фарерского, за ним следует инари-саамский, а остальные используются намного меньше.



Рисуно

к 2: Версии Википедии > 1000 слов в Северной Америке, России и странах Северной Европы



Размер Википедии относительно размера языкового сообщества

Логично было бы предположить, что чем больше носителей языка, тем больше статей на этом языке. Как показано на рисунке 3, такая корреляция действительно существует, но носит слабо выраженный характер ($R^2 = 0,37$, при этом идеальной корреляцией была бы $R^2 = 1,0$). К статистическим выбросам со «слишком малым количеством статей» можно отнести якутский, коми, тувинский и бурятский, а к статистическим выбросам со «слишком большим количеством статей» — фарерский, северносаамский, вепсский и инари-саамский. При этом к статистическим выбросам со слишком малым количеством статей относятся четыре языка в России с крупнейшими языковыми сообществами. Однако по сравнению с более крупными версиями Википедии, не относящимся к приполярному региону, разница между первыми и вторыми не столь уж значительна. Имея показатель в 88 носителей в расчете на одну статью, бурятский находится на одном уровне с албанским (72), а коми и якутский (29 и 34 носителя на статью) сопоставимы с датским (21), венгерским (30) и болгарским (33). Фарерский язык находится на одном уровне с исландским (примерно по 5,5 носителя на статью).

Среди приполярных языков, на которых говорят более 10 000 человек, 5 языков выделяются «слишком большим количеством статей», причем все они находятся в странах Северной Европы или в Карелии: фарерский, северносаамский, вепсский, ливвиковское наречие (олонецкий) и инари-саамский. В случае с самыми маленькими языковыми сообществами (вепсское и инари-саамское) небольшое количество носителей в расчете на одну статью можно было бы назвать вполне логичным. Тем не менее, эти языковые сообщества отличаются более высокой активностью по сравнению с остальными, особенно учитывая их размер.

Таким образом, корреляция между размером языкового сообщества и размером версии Википедии является слабой, и должны быть другие факторы, помимо размера языкового сообщества, определяющие размер версий Википедии.

Авторы

В случае с сообществами столь малого размера, как у приполярных языков, активность отдельных авторов играет более значимую роль, чем в случае с более крупными языковыми сообществами.

Для каждой версии Википедии есть страница в разделе *Специальные страницы*> *Список активных пользователей*, на которой показано, кто проявлял активность в течение последних 30 дней. Термин *Пользователи* Википедии относится к людям или к компьютерным программам (ботам), редактирующим версии Википедии. Некоторые из перечисленных пользователей действительно знают язык рассматриваемой версии Википедии, но значительная часть или даже большинство пользователей этим языком не владеют. Таких лиц, не являющихся носителями языка, можно разделить на три категории: пользователи, которые не касаются лингвистического контента, пользователи, которые в достаточной степени владеют языком и осознают пределы своих возможностей как авторов для написания текста приемлемого качества, и пользователи, которые плохо знают язык или не знают его вовсе, но все равно создают текст, качество которого обычно варьируется от очень низкого до полной тарабарщины.

Чтобы разделить пользователей на носителей и неносителей, принимались во внимание как собственные утверждения лиц о владении языком, так и качество написания текстов. Многие пользователи заявляют о знании языка на своей странице. В рамках исследования утверждения лиц, которые заявляли о своем незнании языка, принимались на веру, а заявления лиц, которые либо утверждали, что знают язык, либо не указывали сведений о том, в какой мере они им владеют, изучались более подробно. Публикации таких пользователей были проанализированы, при этом пользователи с публикациями, которые не требуют языковых навыков, были классифицированы как неносители⁷.

Если взглянуть сначала на группу пользователей, владеющих рассматриваемым языком, то картина выглядит следующим образом: за последний месяц только 3 версии Википедии имели более одного активного пользователя со знанием языка: *якутский* (15 пользователей), *фарерский* (9) и *инари-саамский* (8). В следующих версиях Википедии был один (1) пользователь со знанием языка за тот же период: *северносаамский*, *вепсский*, *коми*, *бурятский*. Остальные версии, содержащие более 1000 слов (*ливвиковское наречие*, *коми-пермяцкий*, *тувинский*, *гренландский*, *инуктитут* и *инупиак*), в течение последних 30 дней не редактировались пользователями со знанием языка.

Если затем взглянуть на совокупность активных пользователей в целом, станет видно, что 1/3 пользователей (44 из 146) проявляют активность более чем в одной версии Википедии на приполярных языках. Это пользователи, которые не знают рассматриваемого языка. До тех пор, пока эти пользователи осознают пределы своих возможностей, их активность, безусловно, приветствуется. Одним из таких примеров является пользователь *Glorious 93* из Франции, который за последний месяц был активен в 8 из 13 версий Википедии на приполярных языках, а также в нескольких версиях в инкубаторе. Активность таких пользователей обычно заключается в добавлении или обновлении ссылок на изображения в информационных блоках — это очень полезная работа без негативных последствий. Представителем второго типа (более) компетентных википедистов является пользователь *Frhdkazan*, который проявляет активность во многих версиях Википедии на языках России.

⁷ Оценка проводилась 6 июня 2021 г.

Согласно его пользовательской странице, его родным языком является татарский, то есть язык тюркской семьи. Языковые версии, на которых он пишет, имеют грамматическую структуру, схожую с татарской. Таким образом, он способен понимать структуру предложений статей, которые редактирует или создает, и при внимательном подходе к редактированию шансы получить текст статьи приемлемого качества весьма высоки.

Третий тип неносителей языка, которые пишут тексты в Википедии, в рамках данной статьи будут именоваться *википедистами-пиратами* (*Wikipedia hijackers*). Такие википедисты обычно выбирают версию Википедии с низкой или полностью отсутствующей активностью, и их публикации быстро занимают доминирующее положение в данной версии Википедии. Одним из таких примеров является версия Википедии на южносаамском языке⁸. Два центральных заголовка («Заглавная страница» и «Добро пожаловать в Википедию, свободную энциклопедию») содержат грубейшие ошибки. Для слова «страница» пользователь использовал слово *bielie* («сторона»), а для слова «свободная» — слово *såångedidh* («делать предложение, просить кого-то стать супругом/супругой»). За этими ошибками скрываются омонимы в норвежском и, следовательно, в норвежско-южносаамском словаре. Норвежское слово *side* означает и «сторона», и «страница», а слово *fri* означает и «свободный», и «делать предложение». Присутствуют также и грамматические ошибки: во фразе *såångedidh daajroegærjase*, которая должна означать «[Добро пожаловать] в свободную энциклопедию», флективная форма *gærjase* (< *gærja*, «книга») неверна; правильный вариант — *gærjese*. Эти фразы даже не взяты со страниц с информационным контентом, они являются частью локализации южносаамской версии Википедии. «Работа» такого типа, возможно, и ведется с наилучшими намерениями, но при этом пользователи совершенно не осознают свои языковые возможности, и это имеет серьезные последствия для данного языкового сообщества. Даже если языковое сообщество не знает о наличии версии Википедии на их родном языке, эти ошибочные формы будут отображаться в результатах поиска в Google, что сделает невозможным использование Интернета для проверки использования языка.

Организованная помощь пользователям: Отделения Викимедиа

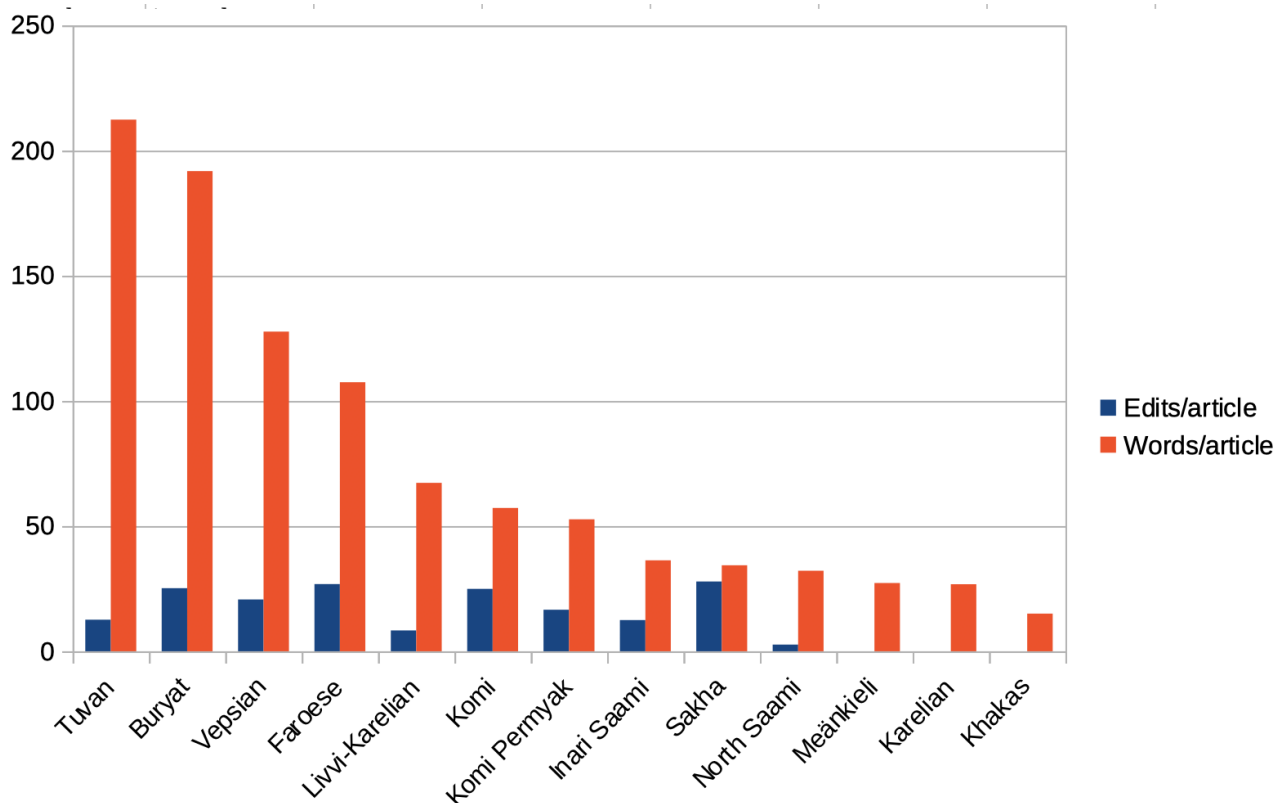
И в России, и в Норвегии, и в Финляндии национальные отделения Викимедиа поддерживают версии Википедии на языках меньшинств соответствующих стран. В России проводились вики-семинары для нескольких языков. Презентацию работы можно найти в публикации Fatkullin 2017. Аналогичным образом, отделения Викимедиа в Норвегии и Финляндии организовали курсы для потенциальных авторов и помогли разобраться с шаблонами, информационными блоками и форматированием.

Длина статьи

Если взглянуть на 12 версий, содержащих более 1000 статей (в данном контексте называемых «большими версиями Википедии»), то только одна из них содержит относительно короткие статьи (версия на хакасском языке, 15 слов на статью), как показано на Рисунке 4. Самые длинные статьи — в версиях на тувинском и бурятском языке (в среднем 212 и 192 слова соответственно).

⁸ <https://incubator.wikimedia.org/wiki/Wp/sma/Aalkoebielie>

32 версии Википедии на приполярных языках содержат менее 1000 статей.⁹; назовем их меньшими версиями Википедии. В таких версиях не только меньше статей, но и размер самих статей тоже меньше. Если взглянуть на каждую версию Википедии по отдельности, то среднее количество слов на статью для более крупных версий составляло 52,7 слова, а для более мелких — 11,5.



Теперь рассмотрим количество правок для каждой статьи. Если взять за основу 321 версию Википедии¹⁰ во всем мире, то они содержат 56 миллионов статей, каждая из которых в среднем редактировалась по 53 раза.

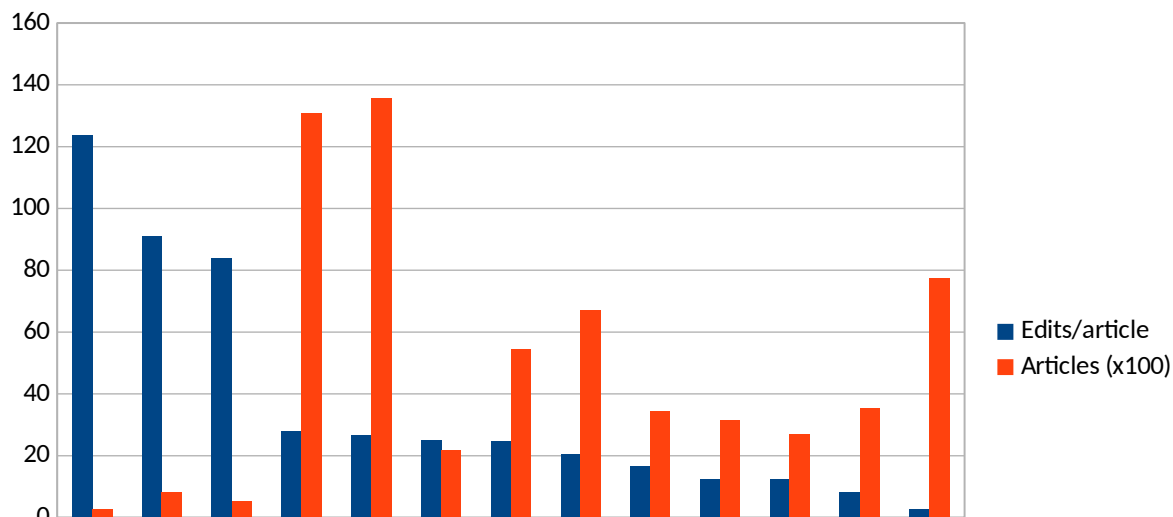
Данные по количеству правок в статьях были доступны только для тех версий Википедии, которые находятся не в инкубаторе. На рисунке 5 показаны версии Википедии на приполярных языках, которых уже вышли из инкубатора. Версии отсортированы по количеству правок на статью, с указанием количества статей.

Как видно из рисунка, корреляции между количеством статей и количеством правок в статьях нет. Рисунок лишь демонстрирует различные версии Википедии. Версии Википедии на гренландском, инуктитуте и инупиаке отличаются очень небольшим количеством статей, но при этом большим количеством правок в каждой статье. Версии на фарерском, якутском и в некоторой степени на коми представляют собой более крупные версии Википедии, в каждой

⁹ Небольшие версии Википедии в порядке убывания (по количеству статей): 100-999 статей: гренландский, ижорский, квенский, инуктитут, чукотский, инупиак, энецкий, центрально-аляскинский юпик, южносаамский, кильдинский саамский и мансийский. 10-99 статей: колтта-саамский, кетский, тундренный ненецкий, тлингитский, лудиковский, луле-саамский, тундренный юкагирский, нанайский, эвенкийский, сибирский юпик, гвичин. 1-9 статей: алеутский, хантыйский, долганский, удэгейский, дене сулин, ороцкий, ульчский, негидальский, тлингитский и хэн.

¹⁰ Обзор версий Википедии см. на <http://wikipedia.org>.

статье которых содержится более 20 правок. Версии же на ливвиковском наречии и северносаамском языке, напротив, содержат очень мало правок в расчете на статью.



Анализ

В ходе анализа будут более подробно рассмотрены различия между версиями Википедии на приполярных языках и изучены причины того, почему одни версии больше других. Также будет обсуждаться вопрос о том, какую роль эти версии могут играть в работе по возрождению языка, не только в контексте практики письма, но и в качестве энциклопедии.

Различия и их причины

Самые развитые версии Википедии можно условно разделить на две категории. Первая категория состоит из версий, которые имеют как высокое соотношение количества статей в расчете на одного носителя, так и относительно большое количество правок на статью. К этой категории относятся две версии Википедии: *Инари-саамская* и *вепсская*. Следовательно, это версии, в которых авторы пишут много статей и уделяют каждой статье какое-то время. В случае с инари-саамской версией на то есть очевидная причина: возрождение инари-саамского языка проводится в плановом порядке Ассоциацией инари-саамского языка, которая ориентируется на отдельно взятое поколение носителей. В настоящее время приоритетом данной организации является повышение грамотности; а одной из стратегий — использование Википедии (см. Olthuis et al 2021, Brecciaroli 2021). Авторы, работающие над вепсской версией Википедии, возможно, взяли за пример финскую и эстонскую версии Википедии.

Ко второй категории относятся версии Википедии с относительно большим населением, поддерживаемые сильной администрацией на родном языке. В эту категорию попадают версии Википедии на *фарерском*, *якутском* языке и *коми*. Возможно, авторы являются участниками успешного движения за возрождение языка и имеют сильную мотивацию, как в случае с авторами статей на инари-саамском, или, может, они принадлежат к языковому сообществу с высоким уровнем грамотности в сфере школьного образования и общества,

охватывающего несколько поколений, как в случае с фарерским языком, якутским языком и коми.

Версии Википедии на *северносаамском языке* и *ливвиковском наречии* не укладываются в такую картину. По сравнению с размером языкового сообщества, они содержат много статей, но в контексте отдельно взятых статей активность невысока. Бурятская и тувинская версии характеризуются большим количеством носителей при относительно небольшом количестве статей. Резкое отличие от ситуации с якутским языком и коми также может носить случайный характер.

Заключение

Главный вывод состоит в том, что очень немногие из версий Википедии на приполярных языках являются активными проектами, курируемыми пользователями, способными писать тексты на рассматриваемом языке. Только в трех версиях Википедии правки вносились более чем одним редактором, действительно владеющим языком, в течение случайного месяца в 2021 году, а еще в четырех версиях правки вносил лишь один редактор, знающий язык. Большинство версий Википедии небольшие, среднее количество статей — 23, и только 13 из 45 версий Википедии на приполярных языках содержат более 1000 статей. Таким образом, большинство версий Википедии создается и наполняется людьми, не знающими язык и не принадлежащими к языковому сообществу.

Успешные версии Википедии на приполярных языках либо опираются на активные и успешные движения за возрождение языков, как в случае с инари-саамским, либо составляют большое сообщество носителей языка с официальной языковой политикой, благоприятствующей развитию приполярных языков и ориентированной на повышение грамотности. Такое наблюдается в случае с якутским, фарерским языком и коми.

Выводы, которые следует сделать для отделений Викимедиа, заключаются в том, что сообщества языков меньшинств не похожи на другие языковые сообщества, и было бы ошибочным ожидать, что они будут вести себя так же, как сообщества языков большинства. Лучшее понимание приполярных языковых сообществ позволит более эффективно предлагать им Википедию, а в некоторых редких случаях даже будет способствовать повышению их грамотности.

Использованная литература

Bilinsky, Yaroslav, 1962: The Soviet Education Laws of 1958-9 and Soviet National Policy. *Soviet Studies* 14:138-157. Oxford

Brecciaroli, Fabrizio 2021: *The Inari Saami Wikipedia*. Wikimedia: Arctic Knot

Fatkullin, Farhad 2017: *Wikipedias in the languages of Russia: in line with the wider Movement strategy?* Poster at the Wikimania 2017 conference. <https://meta.wikimedia.org/wiki/File:WLR-Wikimania2017-poster-frhdkazan.pdf>

Frederiksen, Katti and Carl Christian Olsen 2017: *Det grønlandske sprog i dag*. Rapport over det grønlandske sprog, standpunkt og anbefalinger. Nuuk.

- Isaev, M.I. 1977: *National Languages in the USSR: Problems and Solutions*. Russian original: Resjenie natsional'no-jazykovykh problem v SSSR. Progress Publishers, Moskva.
- Knudsen, Karin Jóhanna L. 2010. Language use and linguistic nationalism in the Faroe Islands, *International Journal of Multilingualism*, 7:2, 128-146, DOI: 10.1080/14790710903518404
- Kreindler, Isabelle (red.) 1985: *Sociolinguistic Perspectives on Soviet National Languages*. Contributions to the Sociology of Language 40. Mouton de Gruyter.
- Meek., A. B. 2010: *We Are Our Language. An Ethnography of Language Revitalization in a Northern Athabaskan Community*. Tucson, Az.: The University of Arizona Press.
- Norwegian Ministry of Labour and Social Inclusion 2009: [Action Plan for Sami Languages](#). Oslo: Norwegian Government Administration Services.
- Olthuis, Marja-Liisa, Suvi Kivelä and Tove Skutnabb-Kangas 2013: *Revitalising Indigenous Languages. How to Recreate a Lost Generation*. Bristol: Multilingual Matters.
- Olthuis, Marja-Liisa, Trond Trosterud, Erika Katjaana Sarivaara, Petter Morottaja and Eljas Niskanen 2021: [Methodological Implications of the Project Čyeti čälled anarâškielân, 'One Hundred Writers for Aanaar Saami'](#) In Indigenous Research Methodologies in Sámi and Global Contexts. New Research - New Voices, Volume: 11
- Pasanen, Annika 2015: *Kuávsui já peeivičuovâ. 'Sarastus ja päivänvalo': Inarinsaamen kielen revitalisaatio*. PhD. dissertation, Helsinki: Helsingin yliopisto.
- Pasanen, Annika 2016: *Saamebarometri 2016 : selvitys saamenkielistä palveluista saamelaisalueella = Sámi giellabaromehter 2016 : čielggadus sámegielat bálvalusain sámeguovllus*. Helsinki: Oikeusministeriö 2016.
- Trosterud, Trond 1995: Sovjetisk språkpolitikk. Finsk-ugriske språk i Russland i eit historisk-politisk perspektiv. *Nordisk Østforum* 3:40-45.

Приложение: Таблицы, показывающие версии Википедии на приполярных языках

Условные обозначения:

Fam = языковая семья (dene = на-дене, ea = эскимосско-алеутская, ie = индоевропейская, mo = монгольская, ps = палеосибирская, tun = тунгусо-маньчжурская, tur = тюркская, ur = уральская).

Reg = регион (a = Северная Америка, n = страны Северной Европы, r = Россия)

Arts = количество статей

Ed/art = количество правок на статью

last30 = активные пользователи (все пользователи), за последние 30 дней

eds l30 = количество правок за последние 30 дней

Spkr = носители языка (согласно данным, указанным на соответствующих страницах на <http://en.wikipedia.org>, главным образом на Ethnologue.

Art/Spkr = количество статей Википедии на носителя

Wrds = общее количество слов в версии Википедии, рассчитанное на основе страниц с контентом из соответствующих дампов Википедии в апреле или мае 2021 г., (как текст между пробелами).

Wrds/art = количество слов на статью

Orth = год введения действующей орфографии, данные главным образом взяты с <http://ru.wikipedia.org>

Версии Википедии на приполярных языках, не находящиеся в инкубаторе

Язык	Fam	Reg	Адрес	Arts	Edits	Ed/art	last30	eds l.30	Spkr	Art/spkr	Wrds	Wrds/art	Orth
инуктитут	ea	a	iu.wikipedia.org	517	43322	83,79	13	36	39475	0,01	6898	13,34	1855
инуитак	ea	a	ik.wikipedia.org	289	35723	123,61	7	11	2144	0,13	1803	6,24	1946
фарерский	ie	n	fo.wikipedia.org	13581	364420	26,83	43	124	72000	0,19	1459615	107,47	1854
северносаамский	ur	n	se.wikipedia.org	7737	20470	2,65	29	75	25000	0,31	248741	32,15	1979
инари-саамский	ur	n	smn.wikipedia.org	2670	33252	12,45	8	959	400	6,68	96942	36,31	1996
гренландский	ea	n	kl.wikipedia.org	810	73832	91,15	15	13	57000	0,01	29683	36,65	1973
якутский	tur	r	sah.wikipedia.org	13077	364499	27,87	39	438	450000	0,03	449228	34,35	1939
вепсский	ur	r	vep.wikipedia.org	6659	137729	20,68	21	487	1600	4,16	850353	127,70	2007
коми	ur	r	kv.wikipedia.org	5441	135522	24,91	21	99	160000	0,03	311417	57,24	1938
ливвиковское наречие карельского языка	ur	r	olo.wikipedia.org	3545	29478	8,32	14	151	31000	0,11	238428	67,26	2007
коми-пермяцкий	ur	r	koi.wikipedia.org	3460	57409	16,59	13	107	63000	0,05	182428	52,72	1938
туванский	tur	r	tyv.wikipedia.org	3167	39925	12,61	9	101	280000	0,01	672601	212,38	1943
бурятский	mo	r	bxr.wikipedia.org	2181	54907	25,18	15	89	265000	0,01	418338	191,81	1939

Версии Википедии на приполярных языках, находящиеся в инкубаторе

Язык	Fam	Cntr	Адрес	Arts	Spkr	Art/spkr	Words	wrds/art	Orth
центрально-алаянский юпик	ea	a	incubator.wikimedia.org/wiki/Wp/esu	193	19750	0,01	1681	8,71	1984
тлингитский	dene	a	incubator.wikimedia.org/wiki/Wp/tli	47	1360	0,03		0,00	1973
гвичин	dene	a	incubator.wikimedia.org/wiki/Wp/gwi/	20	560	0,04		0,00	1970
алеутский	ea	a	incubator.wikimedia.org/wiki/Wp/ale	8	150			0,00	1990
дене сулин	dene	a	incubator.wikimedia.org/wiki/Wp/chp	4	11325	0,00		0,00	1904
хэн	dene	a	incubator.wikimedia.org/wiki/Wp/haa	1	20	0,05	1	1,00	1970
тлингитский	dene	a	incubator.wikimedia.org/wiki/Wp/tli	1	1360	0,00	9	9,00	1973

меянкиели	ur	n	incubator.wikimedia.org/wiki/Wp/fit	1841	30000	0,06	50167	27,25	1985
квенский	ur	n	incubator.wikimedia.org/wiki/Wp/fkv	610	5000	0,12	16191	26,54	2008
южносаамский	ur	n	incubator.wikimedia.org/wiki/Wp/sma	160	600	0,27	1031	6,44	1974
колтта-саамский	ur	n	incubator.wikimedia.org/wiki/Wp/sms	90	320	0,28	1149	12,77	1973
луле-саамский	ur	n	incubator.wikimedia.org/wiki/Wp/smj	43	650	0,07	379	8,81	1983
карельский	ur	r	incubator.wikimedia.org/wiki/Wp/krl	1808	36000	0,05	48383	26,76	2007
хакасский	tur	r	incubator.wikimedia.org/wiki/Wp/kjh	1445	42604	0,03	21745	15,05	1939
ижорский	ur	r	incubator.wikimedia.org/wiki/Wp/izh	677	820	0,83	31391	46,37	2014
чукотский	ps	r	incubator.wikimedia.org/wiki/Wp/ckt	415	5095	0,08	3923	9,45	1937
энецкий	ur	r	incubator.wikimedia.org/wiki/Wp/enf	197	43	4,58	654	3,32	1980
кильдинский									
саамский	ur	r	incubator.wikimedia.org/wiki/Wp/sjd	143	340	0,42	5165	36,12	1985
мансийский	ur	r	incubator.wikimedia.org/wiki/Wp/mns	118	940	0,13	4053	34,35	1938
кетский	ps	r	incubator.wikimedia.org/wiki/Wp/ket	71	213	0,33	734	10,34	1980
тундреный ненецкий	ur	r	incubator.wikimedia.org/wiki/Wp/yrk	66	21926	0,00	3555	53,86	1937
людиковский	ur	r	incubator.wikimedia.org/wiki/Wp/lud	43	3000	0,01	393	9,14	1993
тундреный юкагирский	ps	r	incubator.wikimedia.org/wiki/Wp/ykg	41	370			0,00	1983
нанайский	ps	r	incubator.wikimedia.org/wiki/Wp/gld	40	1400	0,03	1628	40,70	1937
эвенский	tun	r	incubator.wikimedia.org/wiki/Wp/evn	26	26580	0,00	2526	97,15	1937
сибирский юпик	ea	r	incubator.wikimedia.org/wiki/Wp/ess	26	1300	0,02	183	7,04	1937
долганский	tur	r	incubator.wikimedia.org/wiki/Wp/dlg	4	1100	0,00	135	33,75	1961
хантыйский	ur	r	incubator.wikimedia.org/wiki/Wp/kca	4	9600	0,00	159	39,75	1937
удэгейский	tun	r	incubator.wikimedia.org/wiki/Wp/ude	4	100	0,04	133	33,25	1989
орочский	tun	r	incubator.wikimedia.org/wiki/Wp/oac	3	8	0,38	123	41,00	1994
ульчский	tun	r	incubator.wikimedia.org/wiki/Wp/ulc	3	150	0,02	116	38,67	1989
негидальский	ps	r	incubator.wikimedia.org/wiki/Wp/neg	2	75	0,03	117	58,50	1992

Приполярными языками, не имеющими своей версии в Википедии, являются:

атна, алуутик, прибрежный цимшианский, дег ксинак, денаина, ненецкий лесной, гитксан, хайда, инуттитут, инувиалуктун, каска, кокума юкагир, коюкон, кусокуум, нижний танана, наукан юпик, нганасанский, нисгаа, северный слэйви, северный тутчоне, пите-саамский, южный слэйви, южный тутчоне, тличо (догриб), танакросс, уме-саамский и верхний танана.