

Борис Орехов, Кирилл Решетников

ГОСУДАРСТВЕННЫЕ ЯЗЫКИ РОССИИ В ВИКИПЕДИИ

К вопросу о сетевой активности миноритарных
языковых сообществ¹

1. Введение

Современное описание российского общества и составляющих его этносов невозможно без учета того «слепа» с языковой ситуацией, который демонстрируется сетевыми ресурсами. Если принимать во внимание тенденции, наблюдаемые ныне в области распределения информационных потоков, то есть основания предполагать, что активная интернет-коммуникация, а также значительный объем и разнообразие выложенных в Сеть текстов свидетельствуют о социальной престижности и полифункциональности соответствующих языков. Кажется разумным предварительно заключить, что языки, обширно задействованные в Интернете, вовлечены в современные культурные, политические, деловые и бытовые практики. И наоборот, ограниченное или sporadическое использование языка в Интернете может отражать его маргинальное положение. Иными словами, сетевая ипостась любого языка, вероятно, коррелирует с его фактическим современным статусом.

Социолингвисты еще не успели во всей полноте оценить эти тенденции. Национальные сегменты российской Сети, находящиеся за пределами русскоязычного интернет-пространства, до сих пор изучались лишь sporadически, а их сводный общий обзор, тем более детализированный, вообще не производился. Работы о языках народов России в Интернете крайне малочисленны², но и общемировой

¹ Проект выполнен при поддержке Центра исследований Интернета и общества РЭШ.

² Орехов Б. В., Галлямов А. А. Башкирский интернет: лексика и прагматика в количественном аспекте // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог» (Бекасово, 30 мая — 3 июня 2012 г.). Вып. 11 (18): В 2 т. Т. 1: Основная программа конференции. М.: Изд-во РГУ, 2012. С. 502–509; Сахарных Д. М. Удмуртский национальный интернет: проблемы развития коммуникативной среды // Коммуникативная природа человека (Первые Петраковские чтения). Материалы Российской научно-теоретической конференции 31 января 2006 г., УдГУ. Ижевск: Издательство Удмуртского университета, 2006. С. 62–64; Сахарных Д. М. Развитие удмуртского национального интернета: преодоление барьеров // Журналистика Удмуртии: история и современность. К 100-летию удмуртской национальной периодической печати. Ижевск, УдГУ, 2006. С. 177–182.

опыт свидетельствует о том, что исследователи все еще не проявляют достаточного внимания к вопросам сетевого функционирования малых языков. Так, в новейшей книге о социальной истории миноритарных языков Франции только в одной главе, посвященной бретонскому, есть краткое упоминание о наличии википедии³ и двух сотен сайтов на этом языке⁴. В очерках, посвященных другим языкам Франции, фактор сетевого присутствия полностью игнорируется и, таким образом, молчаливо переводится в разряд малозначимых для оценки витальности и распространенности языка. Поэтому проблему аналитической оценки интернет-ресурсов на малых языках можно считать, с одной стороны, назревшей, а с другой стороны — практически неразработанной.

Кроме собственно обозрения соответствующих языковых сегментов Сети, важным был бы также систематический анализ социального и политического аспектов интернет-активности, которую проявляют носители миноритарных языков. Скорее всего, корреляция онлайн-овой и оффлайн-овой версий национального языка не подчиняется абсолютным закономерностям. Дополнительные факторы, корректирующие общую модель корреляции сетевого и внесетевого бытования языка, относятся как раз к области языковой политики и ее социальных следствий. Возможно, и даже скорее всего, есть случаи, когда относительно обширное этническое сообщество, активно использующее национальный язык, слабо распространяет эту активность на Интернет. Если это так, то было бы интересно разобраться в причинах, приводящих к такой ситуации, и сравнить подобные случаи с примерами более симметричного сетевого присутствия. Одним из релевантных факторов здесь вполне может оказаться степень распространенности и доступности Интернета в том или ином регионе. Однако наряду с этим важную роль может играть и политика администрации региона, проводимая по отношению к языку.

2. Википедия и миноритарные языки

Оценка объемов интернета для каждого из языков народов России — дело технологически крайне трудоемкое. Эта задача предполагает доступ к серьезным серверным мощностям, дающим возможность «просеять» содержимое Сети, вычленив и посчитать тексты на языках России. При этом, по данным аналитической компании IDC, в 2009 г. объем информации в Интернете достигал 500 экзобайтов

³ Слово *Википедия* пишется в статье с прописной буквы в том случае, если речь идет об интернет-энциклопедии и соответствующем портале в целом. Строчная буква используется для обозначения языковых разделов. Аналогичный принцип выдержан и при написании слова *Интернет*.

⁴ *Histoire sociale des langues de France / Sous le direction de G. Kremnitz. [s.p.]: Presses Universitaires de Rennes, 2013. P. 450.*

и удваивался каждые полтора года. Ясно, что обойти всю Сеть и проанализировать каждую страницу могут позволить себе только очень крупные компании.

Однако в Интернете присутствует авторитетный ресурс, который, помимо прочих своих свойств, может играть диагностическую роль с точки зрения сетевого присутствия миноритарного языка. Речь идет о свободной интернет-энциклопедии Википедия, где существуют разделы на разных языках мира, в том числе и на языках народов России. Удобство этого ресурса для исследовательских целей в том, что разделы Википедии исчислимы: о них можно говорить не только в терминах бинарных оппозиций (существует — не существует), но и оценивать их количественно — по числу статей, числу участников, числу активных участников и т.д. Имеющаяся статистика открыта и доступна всем желающим.

В то же время к «диагностичности», о которой мы сказали выше, приходится относиться с осторожностью. В Википедии в разное время создавались разделы, которые не могут служить показателем распространенности языка даже в Интернете. Так, история Википедии знает языковые разделы на клингонском, сибирском или старославянском языке (из трех перечисленных только последний раздел существует на момент написания этой статьи), то есть на языках с нулевым числом носителей. Это означает, что сама по себе величина разделов Википедии (объемы которых в случае несуществующих и мертвых языков достигали нескольких сотен статей) как минимум не полностью зависит от числа носителей языка. Однако и в плане корреляции между сетевым присутствием языка вне Википедии и объемом раздела Википедии на этом языке тоже не все гладко. Так, например, в Интернете можно обнаружить и другие сайты на клингонском, включающие разные формы репрезентации языка, в том числе и такие, как общение на форумах⁵. Однако затруднительно найти аналогичные сайты на старославянском, где были бы представлены не только цифровые копии старославянских памятников, но и новые тексты, созданные современными пользователями Интернета по образцу аутентичных.

3. Государственные языки

Статус государственного языка определяется законодательно и поэтому является непосредственным инструментом языковой политики. Государственными называются языки, которым юридически приписывается наивысший ранг, что может проявляться в различных особенностях их функционирования. Так, на государственных языках должно осуществляться делопроизводство; в некоторых случаях, когда

⁵ Например, клингонский форум: URL: <http://www.klingon.org/smboard/> (дата обращения: 15.11.2014).

это установлено законом, на государственном языке должны предоставляться инструкции к потребительским товарам и т.д. Для большинства национальных республик в составе Российской Федерации законодательно установлены свои государственные языки.

Возникает вопрос: способна ли такая функция языковой политики, как установление государственного статуса, повлиять на сетевое бытование языка? В настоящей статье материалом для наблюдений, позволяющих дать предварительный ответ, послужит репрезентация миноритарных государственных языков в Википедии.

Необходимо уточнить, какие именно языки попали в нашу выборку. Во-первых, нас интересовали только языки Российской Федерации. То есть речь идет о языках, являющихся титульными для различных национальных административных образований: о татарском, чувашском, марийском, якутском, чеченском, абхазском и др. Во-вторых, к анализу не привлекаются языки, распространенные на территории РФ, но одновременно являющиеся титульными для других государств (скажем, тексты на украинском или казахском). Мы исходим из того, что создание и размещение текстов на языках этой последней категории осуществляется преимущественно в других странах. В нашем исследовании мы хотели бы ограничиться пределами Российской Федерации, а при анализе данных по этим языкам было бы довольно трудно с уверенностью отличить российскую составляющую от зарубежной. С одной стороны, такой подход кажется естественным. С другой стороны, он приводит к неочевидным решениям. Так, например, из нашей выборки исключен осетинский язык, имеющий государственный статус на территории одного из субъектов РФ — Республики Северная Осетия — Алания, поскольку одновременно осетинский является и титульным языком другого (частично признанного) государства — Южной Осетии.

Таким образом, в сферу нашего рассмотрения попали следующие языки: *абазинский, аварский, агульский, адыгейский, акушинский* (диалект *даргинского*, на основе которого был создан литературный язык), *башкирский, бурятский, горномарийский, ингушский, калмыцкий, карачаево-балкарский, коми-зырянский, кумыкский, лакский, лезгинский, лугово-восточный марийский, мокша-мордовский, ногайский, рутульский, табасаранский, татарский, татский, тувинский, удмуртский, хакасский, цахурский, чеченский, чувашский, эрзя-мордовский, южно-алтайский, якутский.*

4. Статус языка и Википедия с точки зрения количественных данных

Не все государственные и титульные языки имеют свои разделы в Википедии. Из числа языков РФ, обладающих статусом государственных (наша выборка, составленная с учетом вышеописанного принципа, включает 31 язык), только 18 демонстрируют в той или

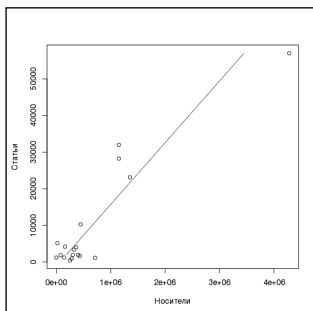
иной мере развитые сегменты в рамках этого энциклопедического ресурса. Своих полноценных разделов не имеют *цахурский, агульский, рутульский, абазинский, хакасский, ногайский, адыгейский, табасаранский, кумыкский, даргинский, татский, южноалтайский, ингушский*. Особенно примечательны в этом списке южноалтайский, карельский, хакасский и ингушский, являющиеся не просто государственными, но и титульными для своих субъектов федерации. И если число носителей карельского и хакасского действительно мало и фиксируется ниже психологически значимого порога в 50 000 человек, а южноалтайский лишь ненамного превышает этот порог, то для ингушского (более 300 тыс. носителей) это скорее странный результат — особенно если учесть существование полноценного раздела для лакского, степень распространённости которого вдвое меньше (145 895 человек по данным переписи 2010 г.). Тем не менее для всех этих языков мы можем обнаружить разделы в так называемом «инкубаторе», т.е. в специальной тестовой зоне, где новые википедии проходят своего рода «испытание»: в том случае, если им удастся получить достаточную поддержку сообщества, они могут превратиться в полноценные разделы. Большинство «инкубаторных» википедий созданы автоматически и не имеют наполнения за пределами статистической погрешности.

Скорее всего, отсутствие полноценных энциклопедических разделов свидетельствует о низкой престижности упомянутых языков даже в тех регионах, которые должны быть базой для их развития и дальнейшего распространения. В то же время на территории России встречаются и языки, не обладающие официальным статусом, но представленные в Википедии полноценным самостоятельным разделом: *коми-пермяцкий, кабардино-черкесский и цыганский*. Это означает, что придание языку государственного статуса само по себе не является достаточным условием для повышения его престижности и активизации его распространения в Интернете.

Количественный анализ показывает, что между численностью говорящих на государственном языке и количеством статей в Википедии нет однозначной зависимости. Хотя на общем фоне коэффициент корреляции Пирсона между числом носителей и количеством статей в Википедии выглядит высоким и составляет 0,922 (что укладывается в доверительный интервал между 0,8 и 0,97), этой тенденции подчиняются не все языки (см. *Илл. 1*). Особенно репрезентативны в этом смысле случаи, когда языки с меньшим числом говорящих обгоняют более распространённые языки по количеству статей. Так, по состоянию на март 2014 г. горномарийская википедия насчитывает 5119 статей, а лугомарийская — 4007 статей, в то время как, по данным Всероссийской переписи населения 2010 г., число носителей горномарийского языка составляет 23 062, а число носителей лугового марийского — 365 316 человек.

Такое положение дел связано с тем обстоятельством, что большой процент статей в интернет-энциклопедии не пишется вручную, а создается автоматически с соревновательными целями и с расчетом на повышение престижности языка (или по крайней мере соответствующего языкового раздела). Нам уже случалось подробно писать об этом феномене роботизированного заполнения Википедии⁶. Его следствием является создание большого числа шаблонных статей, по многим параметрам не отражающих языковую реальность.

Иллюстрация 1. Зависимость числа статей в Википедии от числа носителей языка



На иллюстрации 1 показан график, на котором точками отмечены реальные показатели числа носителей (по оси x) и показатели количества статей в соответствующих википедиях (по оси y). Кроме того, на графике изображена прямая, отражающая ожидаемые значения линейной регрессии.

Интерпретировать этот график следует таким образом: языки, попавшие в правую его часть (т.е. те, которые отмечены справа от прямой модельных значений), демонстрируют слабый интерес носителей языка к созданию википедии. В упрощенном виде это формулируется так: *носителей много, статей в википедии мало*. Среди языков этой группы — татарский и аварский.

Показатели слева от прямой демонстрируют большой интерес интернет-активистов к наполнению википедии, чем это можно было бы ожидать, исходя из числа носителей (*носителей мало, статей в википедии много*). К языкам этой группы относятся башкирский, чувашский, якутский, горномарийский.

Почти соответствуют модельному значению реальные показатели чеченского языка.

⁶ Орехов Б.В., Решетников К.Ю. К оценке Википедии как лингвистического источника: сравнительное исследование // Современный русский язык в интернете. М.: Языки славянской культуры, 2014. С. 310–321.

Наибольшее положительное отклонение истинного значения от модельного составляет 831 990, наибольшее отрицательное — -815 218. Среднеквадратичное отклонение для данной модели составляет 0,8418.

Помимо прочего эта статистика говорит о том, что, несмотря на автоматическое преумножение статей, т.е. наличие искусственного фактора, обуславливающего увеличение их числа, существуют и определенные рамки, в которых остаются активисты национальных википедий. Раздел Википедии на языке x скорее всего не превысит по объему раздел на языке y , если число носителей языка y на порядок больше, чем у языка x (скорее всего, но не обязательно: число носителей горномарийского именно что на порядок меньше, чем у лугового марийского).

5. Судьба миноритарных языков как предмет дискуссии: функционализм vs. «экологизм»

Обзор различных точек зрения на репрезентацию малого языка в Интернете показывает, что по крайней мере в сетевых дискуссиях эти точки зрения чаще всего высказываются именно в связи с Википедией.

В 2010 г. Институтом повышения квалификации и переподготовки работников образования Удмуртской Республики был получен грант Общества М.А. Кастрена (Финляндия) на нужды развития удмуртского раздела Википедии. Это событие послужило поводом для дискуссии о необходимости сохранения малых языков, включая частный аспект данной проблемы: необходимость их сетевого существования.

Пользователь StopKran на известном ресурсе «Хабрахабр» высказался следующим образом:

Я думаю что если язык умирает то это не просто так. Язык это в первую очередь средство донесения информации, и чем удобнее язык тем лучше он выполняет свою функцию... Английский выполняет функцию языка гораздо лучше, многих других. Худшие выбывают, лучшие остаются на местах. Я конечно не знаю удмуртский, но я не думаю что это какой-то мега революционный язык. Поэтому думаю лишись его будет только лучше. Надо идти вперед и развивать языки, а не оглядываться назад и сохранять старое⁷.

Действительно, существует точка зрения (немаловажная для анализа интернет-присутствия и сетевой поддержки языков), согласно которой языковое многообразие, наблюдаемое на данный момент в мире, и в частности в России, избыточно: оно не требуется для выполнения тех функциональных задач, которые должен выполнять

⁷ Цит. по: http://habrahabr.ru/post/91454/#comment_2759582 (дата обращения: 10.11.2014). Орфография и пунктуация цитируемых источников сохранены.

язык. Язык, как считают сторонники этого мнения, нужен только для передачи информации и для интерактивного коммуникационного взаимодействия. Строго говоря, естественный язык — не единственный способ, с помощью которого можно передавать информацию, хотя именно он на данный момент и последние 40 000 лет является самым востребованным для этого инструментом. С функциональной точки зрения пестрота языковой карты даже затрудняет информационный обмен, особенно в том случае, если участники коммуникативной ситуации владеют непересекающимися множествами языков. Такая точка зрения сформировалась не только благодаря своей (кажущейся) логичности, но и благодаря хорошо известным унифицированным моделям коммуникации, которые функционируют в больших регионах, обслуживаемых преимущественно одним языком. На всех соответствующих территориях распространены и малые языки, но в качестве *lingua franca* используется один «большой» язык, которого бывает достаточно для комфортного существования человека, перемещающегося в пространстве внутри такого региона. Как известно, «большим» языком стал, в частности, английский — прежде всего в Северной Америке, но также, пусть и со значительными оговорками, в большей части остального мира. Специалисты знают, что в Северной Америке распространены различные языки (не только индейские, но и языки диаспор), однако очевидно и то, что конвенциональный язык этого региона — английский. Похожим образом обстоит дело и с русским на территории бывшего СССР, и с испанским в Латинской Америке, а отчасти и с китайским в некоторых странах Юго-Восточной Азии. Есть и менее распространенные языки, играющие сходную роль в пределах определенных ареалов. Так, в меньшем масштабе аналогом русского окажется, скажем, суахили. Подобные примеры рождают и поддерживают представление о том, что моноязычную модель коммуникации можно глобализировать и окончательно распространить на весь мир, нивелировав языковое разнообразие, сведя его к минимуму, в идеале — к одному-единственному, общему для всех языку. Это представление может сыграть свою роль в случае, когда оно присуще ответственным лицам, которые распределяют материальные ресурсы, или интернет-активистам, которые создают и поддерживают технические платформы для коммуникации и распространения контента в Сети. Скажем, если бы адептами этого взгляда были представители фонда Викимедиа, занимающегося поддержкой Википедии, то настоящий обзор сетевой репрезентации малых языков был бы невозможен.

Вопрос о том, можем ли мы заботиться об одном языке и жертвовать остальными, исходя из того, что функциональность общения от этой жертвы не пострадает, переносится в философскую, культурно-историческую и этическую сферы. На этот вопрос, очевидно, возможны два противоположных ответа. Так, участники недавней

международной конференции по проблемам ревитализации языков, находящихся на грани исчезновения⁸, единодушно исходили из характерного «экологического» взгляда, согласно которому все языки (и «большие», и малые) следует считать одинаково ценными. Однако как видно, например, по репликам пользователя StopKran и его сторонников в дискуссии на «Хабрахабре», противоположное мнение также распространено.

Убежденность в необходимости поддерживать малые языки отчасти связана с сугубо научными проблемами, волнующими специалистов. Например, исчезновение даже одного языка приведет к потере важного источника информации о лингвистической и этнической истории, а также о разного рода лексических, грамматических и фонетических явлениях. Однако необходимо отметить, что идеология сохранения языков сопряжена также с политическими установками мультикультурализма⁹, которые, в частности, предполагают противодействие дискриминации меньшинств. Одна из ключевых мультикультуралистских идей состоит в том, что любой человек должен иметь выбор (например, выбор в отношении того, каким языком пользоваться), а представители миноритарных сообществ не должны сталкиваться с ситуацией, при которой следующее поколение вынуждено избирать для общения доминирующий на данной территории язык без возможности использовать другие. Вынужденный переход на доминирующий язык происходит, например, тогда, когда малый язык игнорируется институтами образования и выбор в пользу малого языка может повлечь за собой отсутствие доступа к социальным лифтам. Такие случаи расцениваются сторонниками мультикультурализма как проявление институционализированной дискриминации, граничащей с культурным фашизмом.

Разумеется, отдельной оговорки требует такая категория, как умирающий язык, число носителей которого не превышает пары десятков человек. Такой язык обречен на гибель в течение одного-двух десятилетий, и этому не смогут противостоять никакие политические усилия. Среди языков народов России в этом положении оказались колтта-саамский, алыторский, энецкий, чулымский, алеутский, орокский, южноюкагирский (колымский, он же одульский), науканский эскимосский. Число носителей этих языков ограничивается пятью десятками. Однако если язык является родным хотя бы для нескольких

⁸ «Language revitalization in a Russian and European context: Exploring solutions for minority language maintenance», Хельсинки, 16–18 декабря 2013 г.

⁹ См., например: Куропятник А.И. Мультикультурализм: проблемы социальной стабильности полиэтнических обществ. СПб.: Изд-во СПбГУ, 2000; Бенхабиб С. Притязания культуры. Равенство и разнообразие в глобальную эру. М.: Логос, 2003; Kymlicka W. Politics in vernacular: Nationalism, Multiculturalism and Citizenship. Oxford: Oxford University Press, 2001.

тысяч человек, то, с точки зрения мультикультурализма, следует прилагать все усилия, чтобы сообщество в языковом смысле выжило.

Ключевую роль в общей системе языковых статусов играет фактор престижности. В мире имеется целый класс миноритарных языков, которые не находятся на терминальной стадии существования: ими пользуются в быту, на них вещают СМИ, создается художественная литература. Но общение на этих языках не считается престижным: сами носители ощущают свой язык как экономически, социально и политически неперспективный и в результате дают новому поколению закономерную установку ориентироваться на «большой» язык. Ощущение непрестижности — важнейший триггер, запускающий процесс исчезновения языка. Именно в связи с этой проблемой и в контексте борьбы с ней Википедия берет на себя несколько неожиданную функцию, становясь каналом трансляции миноритарных языков и культур, инструментом для их «рекламы» и продвижения.

б. Википедия как канал трансляции культуры

Как можно судить по сетевым высказываниям представителей миноритарных языковых сообществ, Википедия является для них престижной формой репрезентации языка.

В октябре 2011 г. на форуме sharvili.com участник под ником Namik555 оставил следующую весьма показательную реплику:

В других википедиях например английской мои попытки улучшить статью «лезгистан» привело тому что какойто ц'апаш начел вести со мной войну правок пастаяна удаляя загруженные мной файлы и тексты. они суют нос во все статьи касающихся лезгин на всех языках. но мы можем создать лезгинскую википедии где мы будем писать с нашей точки зрения а не то что нам суют¹⁰.

Как нам представляется, в этом высказывании отражены два взаимосвязанных обстоятельства:

1. На международных ресурсах у представителей малых народностей возникают проблемы с распространением информации о своем языке и о своей культуре («они суют нос во все статьи касающихся лезгин на всех языках»).

2. У таких языков, как лезгинский, нет достаточного числа площадок для продвижения, а Википедия рассматривается именно как перспективная площадка, притом уже заранее обладающая необходимым ресурсным весом («мы можем создать лезгинскую википедии

¹⁰ Цит. по: http://sharvili.com/forum/thread/42494/википедия-на-лезгинском-языке/view_54893/ (дата обращения: 09.10.2014).

где мы будем писать с нашей точки зрения»). Этих двух пунктов мы последовательно и коснемся.

Необходимо учитывать, что деятельность по созданию национальных языковых ресурсов осуществляется в условиях доминирования «больших» культур и языков (английского в глобальном контексте, русского — на постсоветском пространстве и т.д.). Господство последних отражается на потенциале и статусе интернет-сообществ: англоязычное культурное пространство, точно так же как и русскоязычное, в своих сетевых сегментах определяет тематику и формы коммуникации. В этих условиях для представителей региональных и малых культур одна из перспективных стратегий заключается в том, чтобы продвигать национальный контент через уже существующие каналы трансляции, которые базируются на «больших» языках и культурах. Именно в рамках такой стратегии пользователем Namik555 была сделана попытка отредактировать статью «Лезгистан» в англоязычной википедии.

У такого подхода есть свои плюсы и свои минусы. Посредством перевода представитель миноритарного сообщества может сообщить нечто о своей культуре большому количеству людей, что в плане трансляции, несомненно, представляет собой выигрышную стратегию. Заметим, что под переводом здесь имеется в виду как языковой перевод, так и перевод в рамках межкультурной коммуникации, при котором осуществляется перекодирование сообщения на язык большой культуры с учетом ее стандартов. Однако на этом пути носитель малой культуры сталкивается с ограничениями двух типов. Во-первых, общим местом в теории межкультурной коммуникации стал тезис о непереводаемости части сообщения в силу объективных факторов (таких, как специфика выразительных возможностей языка, реалии и проч.). Во-вторых, текст, созданный носителем малой культуры и представленный на площадке, для которой базовой является большая культура, может быть отредактирован носителем этой большой культуры. В ситуации, когда на англоязычном ресурсе страницу контролирует модератор или редактор, приоритет естественным образом остается за ним. Даже если представитель малой культуры (например, лезгин) пишет на хорошо знакомую ему тему и сам является источником информации, приоритетных прав на аутентичное высказывание в таких условиях у него нет. Соответственно, какие-то компоненты первоначального сообщения могут быть учтены и восприняты, а какие-то подвергнуты цензуре. В итоге передаваемый месседж должен будет видоизмениться, и получившийся результат может не устроить отправителя сообщения. Так произошло с пользователем Namik555, который, насколько можно судить, воспринял действия модераторов как своеобразный выпад против всего представляемого им народа. Таковы вкратце сильные и слабые стороны «переводной» стратегии.

Другой способ состоит в том, чтобы продвигать и развивать собственную культуру (малую или региональную) в рамках соответствующего языкового пространства — например, создавать коммуникативные площадки, предполагающие общение на конкретном миноритарном языке. При этом Википедия как интернет-проект, обладающий ресурсным весом и массово посещаемый, заведомо будет иметь преимущество перед большинством остальных возможных вариантов, становясь приоритетным предметом рассмотрения при выборе сетевого плацдарма. Предназначение Википедии с точки зрения первоначального замысла — агрегирование и свободное распространение накопленного человечеством знания. Википедия должна выполнять справочные функции. Однако в условиях культурного неравноправия для представителей малых народов этот интернет-портал становится не в последнюю очередь местом реализации языковых амбиций.

Успешность применения этой «аутентичной» стратегии в значительной степени зависит от наличия материальных ресурсов. Весьма сложно без специальной поддержки сделать из интернет-сайта авторитетную коммуникативную площадку, роль которой в продвижении культуры была бы очевидной и не сводилась бы к маргинальной инициативе частного характера или к интересам академического круга. Трудность выполнения такой задачи в современных условиях обычно связана с полной сторонней занятостью потенциальных участников проекта. В текущей ситуации ощутимый положительный эффект могут обеспечить только два фактора: личный энтузиазм активистов (число которых должно быть если не большим, то все же не ничтожно малым) и использование внешних ресурсов поддержки, чаще всего связанных с возможностями государственных институтов. Эти институты способны предоставить прежде всего необходимую внесетевую базу для интернет-присутствия языка. Крайне сомнительны перспективы развития форумов, чатов, социальных медиа и прочих площадок для сетевого общения на том или ином языке, если на этом языке нет школьного образования, не функционируют СМИ, не развивается художественная литература. Поэтому, несмотря на то что Википедия рассматривается нами как индикатор сетевой представленности языка, мы не можем изучать эту форму репрезентации отдельно от внесетевых социолингвистических факторов. Иными словами, успех ретрансляции малых культур в Сети напрямую зависит от механизмов лоббирования этнонациональных интересов в госструктурах.

В этом смысле показательно, что упоминавшийся выше грант, полученный одним из педагогических институтов Удмуртии, был выдан именно на разноплановую деятельность по развитию удмуртской википедии, а не на написание конкретных статей (так как наполнение Википедии должно производиться силами энтузиастов).

С учетом этих оговорок представляется актуальным поставить следующий вопрос: перспективны ли ресурсы типа Википедии (и прежде всего сама Википедия) на миноритарных национальных языках в качестве провайдеров языковой и культурной независимости для соответствующих сообществ? Является ли работа над такими сайтами всего лишь примером наивного национального романтизма в применении к Интернету, или же она создает предпосылки для каких-то серьезных, качественно новых практик? Имеем ли мы дело с поверхностным прожектерством — или с появлением точки роста, благодаря которой Интернет в ближайшем будущем может обогатиться жизнеспособным альтернативным контентом?

7. Две опасности на пути создания национальных википедий

Чтобы ответить на этот вопрос, нужно оценить реальную практику работы над википедиями на малых языках РФ. Эта практика по большей части порождает тексты, с лингвистической точки зрения не вполне аутентичные и отражающие не столько реальный языковой узор, сколько случайные и довольно специфические фрагменты языка, связанные с определенными темами. Например, в мокша-мордовской википедии обнаруживает себя то, что мы называем «синдромом орхидеи». Из-за роботизированного наполнения интернет-энциклопедии последняя оказывается крайне диспропорциональной в лексико-тематическом плане; в частности, слово «орхидея» в ней встречается гораздо чаще, чем слова, которые мы, исходя из наших знаний о типологии естественных языков, ожидали бы увидеть на верхних строчках частотного списка, составленного по текстам википедии (топ частотных списков в любом языке обычно состоит из служебных слов). Так происходит потому, что разновидностям орхидеи посвящено огромное в долевым отношении число статей.

Можно сказать, что в сфере сетевой репрезентации малого языка на базе Википедии обнаруживаются как минимум две опасные перспективы.

Первая связана с возможными проявлениями идиостилистического волюнтаризма. В языке тех википедий, которые пишутся на некоторых не самых распространенных языках, мы находим сущности, отсутствующие в речевой практике носителей. Эта проблема имеет причиной искусственность большинства литературных вариантов малых языков. Для понятий и реалий, обычно не затрагиваемых в разговорной речи или же обозначаемых в бытовом узусе русскими заимствованиями, в литературном языке создаются специальные неологизмы. Например, в текстах, отражающих литературную разновидность коми языка и посвященных специфическим темам (в частности, сфере государственного устройства, социальных услуг и т.д.),

встречается слово *каналан* ‘государственный’, активно используемое и в соответствующей википедии: «Удмурт кыв (удмурт кыл) тайӧ Удмуртия Республикаын каналан кыв. Сійӧн сернитӧ Удмуртия Республикаса вужвойтыр — удмурт йӧз (матӧ 484 сюрс морт)» (статья «Удмурт кыв»); «Коми Республикалӧн дӧрапас — Коми Республикалӧн каналан символгясысь ӧти» (статья «Коми республикаса дӧрапас»). Сюда же могут быть отнесены слова *кансям* ‘политика’, *оланподув* ‘конституция’, *мусертас* ‘карта, очертания земли’ и т.д. В ходе полевой лингвистической работы было установлено, что по крайней мере носителям ижемского диалекта эти слова неизвестны и не опознаются ими как аутентичные для коми языка¹¹. Обычно слова такого рода свойственны исключительно письменной речи узкого круга журналистов (а также устной речи журналистов местных телеканалов), а в живом общении на месте этих слов используются русские заимствования. Такая ситуация в области терминологической лексики связана с тем, что сфера функционирования коми, как и других малых языков, является довольно узкой. Это явление отражается и при написании статей для википедий. Вполне вероятно (и даже наблюдается в действительности) положение, при котором вариант языка, представленный в интернет-энциклопедии, не используются вне этого жанрового поля. Кроме коми, мордовских и других википедий, созданных на государственных языках РФ, в связи с данным феноменом можно упомянуть также цыганскую википедию. В силу своей диалектной принадлежности язык этой википедии, теоретически говоря, может быть отчасти понятен одной из групп цыган, присутствующих на территории России, а именно группе, обозначаемой этнонимом «кэлдэрары». Однако есть небезосновательные сомнения относительно того, что этот язык может выглядеть как нечто аутентичное и опознаваемое для других ветвей цыганского этноса, в частности для русских цыган — носителей диалекта, известного как *русска рома*¹².

Иными словами, проекты, подобные тому, который предлагает в вышеприведенной цитате пользователь Namik555, чреваты возникновением искусственных лингвистических конструктов, непонятных большинству носителей языка и не отражающих реальную практику его употребления. Национальный языковой энтузиазм может приводить к культивированию фиктивных сущностей, создаваемых ограниченным числом людей и незнакомых остальным представителям этноса (заметим, что обычно энтузиастов в деле поддержки и расширения малых языков немного).

¹¹ Авторы благодарят Е.В. Кашкина за консультацию и предоставленный языковой материал.

¹² Авторы благодарят М.В. Ослона за консультацию по этому вопросу.

Помимо неумеренного использования «авторской» лексики, существует и еще одна опасность. В целях конкуренции с другими, «соседними» миноритарными разделами Википедии создатели статей зачастую прибегают к таким способам пополнения контента, которые негативно отражаются на качестве и содержательной ценности ресурса в целом. Речь идет об автоматически генерируемых статьях, или о так называемых «ботозаливках», т.е. о «заливках», совершаемых роботами («ботами»). Результатом обычно становятся текстовые заготовки, созданные по одному шаблону, в которые механически добавляются данные, легко извлекаемые из того или иного внешнего источника. Одним из таких источников служит, например, Реестр водных ресурсов России, откуда авторы башкирской и татарской википедий берут информацию для тысяч статей о реках, протекающих на территории РФ. Фактически мы наблюдаем парадокс: с одной стороны, энциклопедия в данном случае пишется в первую очередь для того, чтобы появился еще один перспективный канал трансляции миноритарного языка и региональной культуры, с другой стороны – качество распространяемого контента сильно занижается уже на раннем этапе развития ресурса. А ведь для эффективного продвижения миноритарных языков чрезвычайно важно, чтобы распространяемый в рамках этой деятельности контент был качественным. В частности, он не должен расширяться за счет материала, который, во-первых, делает содержательное наполнение ресурса диспропорциональным, а во-вторых, опять-таки воспринимается носителями языка как нечто не (вполне) аутентичное. Между тем «ботозаливки», конечно, порождают именно такой сомнительный материал¹³.

Востребованность «ботозаливок» связана с тем, что одновременно со стратегической задачей трансляции языка и культуры решается также другая, тактическая и, по сути, сиюминутная задача. Речь о таком тривиальном и характерном для сетевой деятельности феномене, как искусственное повышение рейтинга. Авторы стремятся поднять «свою» википедию как можно выше в перечне разделов, ранжированных по числу статей, и в итоге внимание фокусируется на количественном критерии, а не на сущностном вопросе о престижности и жизнеспособности малых культур. Автоматическое увеличение объема во многом обесмысливает все предпринимаемые усилия, и социально активный представитель миноритарного этноса вправе задать вопрос: зачем моему языку и моей культуре ресурс, состоящий по большей части из того, что не отражает ни язык, ни культуру?

¹³ Подробнее об этом см.: *Орехов Б.В., Решетников К.Ю.* К оценке Википедии как лингвистического источника: сравнительное исследование // Современный русский язык в интернете. М.: Языки славянской культуры, 2014. С. 310–321.

К этому нужно добавить, что «ботозаливки» на самом деле слабо влияют на статус языкового раздела, почти не повышая его ценность ни в глазах заинтересованной «википедийной» аудитории, ни за ее пределами. Таким образом, учитывая не только качественные потери, но и отсутствие ощутимого результата, тенденцию к автоматическому увеличению объема миноритарной википедии следует признать тупиковой.

Необходимо отметить также, что само по себе использование национальной википедии в качестве инструмента для укрепления языковых и культурных позиций выглядит довольно специфично. Собственно, здесь мы опять-таки имеем дело с некоторым противоречием. Как было замечено выше, Википедия, вообще-то, создана для справочных целей: она должна функционировать как хранилище и бесплатный источник всевозможной информации — научного знания, биографических данных, исторических сведений и т.д. Вопрос о том, на каком языке она написана, с точки зрения ее энциклопедической функции вторичен: в идеале национальная википедия — это всего лишь версия единого информационного ресурса, создание которой носит сугубо прагматический характер. Такая логика плохо сочетается с идеологией этнонационального «культуртрегерства», о которой говорилось в этом и предыдущем разделах. Будучи поставлена на службу совершенно особым локальным целям, а именно целям позиционирования и укрепления миноритарной идентичности, Википедия, строго говоря, перестает выполнять свою первоначальную задачу.

Другое дело, что не имеет смысла возражать против определенной тематической и структурной специфики, характеризующей практически любую языковую версию Википедии, включая также и те, которыми представлены крупные языки и культуры. Раздел, создаваемый на языке той или иной страны, содержит, как правило, статьи о местных и локально значимых реалиях, которые в других википедиях либо не освещаются вообще, либо освещаются гораздо более скупо. В этом смысле разные национальные разделы Википедии в какой-то степени дополняют друг друга, расширяя, каждый со своей стороны, территорию всеобщего знания. Однако специфические сведения регионального или этнонационального характера вряд ли могут играть главенствующую роль в энциклопедическом источнике — с точки зрения функциональной ценности для него предпочтителен, вероятно, определенный количественный баланс между разными категориями информации. В принципе такого баланса мы вправе ожидать и от миноритарных википедий. Наличие качественных статей, посвященных темам из области универсального знания, может оказаться не менее, если не более существенным для востребованности и положительной оценки миноритарного языкового раздела, нежели освещение местных культурных или географических реалий.

Но как бы то ни было, вышеописанные «культуртрегерские» и соревновательные тенденции в настоящий момент остаются характерными для развития википедий на малых языках РФ, и если ситуация изменится, то это едва ли произойдет благодаря чьим-либо академическим рекомендациям.

8. Заключение

Думается, что решение проблемы сетевого функционирования малых языков, рассмотренной нами на примере Википедии, в целом зависит от нескольких факторов. Все они были нами упомянуты; в порядке предварительного вывода мы выскажем рабочую гипотезу, в рамках которой попробуем ранжировать эти факторы по степени значимости.

Для (успешного) существования сетевой версии малого или регионального языка важны как минимум следующие параметры:

1) наличие или отсутствие государственного или иного официального статуса;

2) количество носителей;

3) степень активности и количество энтузиастов, использующих и продвигающих язык в Сети;

4) более или менее интенсивная культурно-образовательная, а также — в первую очередь — спонсорская и организационная деятельность государства и различных заинтересованных учреждений, как государственных, так и частных, направленная на создание благоприятных внесетевых условий для бытования языка.

В настоящий момент, учитывая проанализированные нами данные по сетевой репрезентации малых языков РФ, мы склонны считать, что релевантность этих параметров возрастает по мере продвижения от начала к концу представленного здесь перечня.

Государственный статус, как мы могли видеть по крайней мере на материале Википедии, не является решающим фактором для «хорошего самочувствия» языка в Сети. Некоторые языки, являющиеся титульными в соответствующих административных образованиях, не имеют собственных энциклопедических разделов (так обстоит дело, к примеру, с ингушским), и наоборот, интернет-присутствие, распространяющееся на Википедию, отмечено для нескольких языков, у которых нет государственного статуса (в частности, для коми-пермяцкого и кабардино-черкесского).

Более существенным выглядит количество носителей: этот параметр дает довольно отчетливую корреляцию между сетевыми и не-сетевыми характеристиками языка, однако рассмотренное множество реальных соотношений не сводится к простой линейной зависимости — напротив, встречаются разнообразные и довольно любопытные отклонения от нее (см. раздел 4).

Наконец, факторы, помещенные в вышеприведенном перечне на третье и четвертое место, можно гипотетически оценить как еще более релевантные, чем количество носителей. В пользу такого предположения говорят, в частности, как раз отступления от линейной закономерности, описанные в разделе «Статус языка и Википедия с точки зрения количественных данных».

Нужно, однако, оговориться, что ранжировать в плане значимости параметры 3) и 4) по отношению друг к другу мы не беремся. Ясно, что существует целенаправленная проектная активность энтузиастов и их партнеров, в рамках которой можно выделить, во-первых, индивидуальные и групповые инициативы по созданию сетевого контента, во-вторых — материальную и координационную поддержку со стороны организаций, прежде всего государственных. Благодаря последней становится возможным, в частности, школьное образование на миноритарном языке, а также повышается его престиж. Если забота о языке превращается в приоритет местной культурной и образовательной политики, то относительно хорошие результаты могут фиксироваться даже для языка с не очень большим числом носителей (как это наблюдается, скажем, в случае якутского). Однако отдать пальму первенства (государственным) кураторам или же, напротив, сетевым активистам в настоящий момент затруднительно.

Что же касается использования Википедии как (единственного) надежного плацдарма для поддержки языка в Интернете, а также применения или неприменения малоэффективных методов типа «ботозаливок», то подобные локальные феномены скорее являются вторичными по отношению к уже сформировавшейся ситуации, нежели обуславливают ее. Радикальная ставка на Википедию и сопутствующие такому подходу авантюрные «технологические» практики могут рассматриваться как признак низкого фактического статуса языка или по меньшей мере как свидетельство слабости его сетевых позиций, т.е. как следствие переживаемого языком кризиса, который вызван «оффлайнowymi» социолингвистическими причинами.

Скажем еще раз, что все эти выводы носят предварительный и гипотетический характер — в частности, потому, что базируются в основном на изучении Википедии, диагностическая ценность которой ограничена. Обработка больших массивов сетевого контента, отражающего использование малых языков, позволит верифицировать сформулированные здесь гипотезы.

Литература

1. Бенхабиб С. Притязания культуры. Равенство и разнообразие в глобальную эру. М.: Логос, 2003.
2. Куропятник А.И. Мультикультурализм: проблемы социальной стабильности полиэтнических обществ. СПб.: Изд-во СПбГУ, 2000.

3. Орехов Б.В., Галлямов А.А. Башкирский интернет: лексика и прагматика в количественном аспекте // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог» (Бекасово, 30 мая — 3 июня 2012 г.). Вып. 11 (18): В 2 т. Т. 1: Основная программа конференции. М.: Изд-во РГГУ, 2012. С. 502–509.
4. Орехов Б.В., Решетников К.Ю. К оценке Википедии как лингвистического источника: сравнительное исследование // Современный русский язык в интернете. М.: Языки славянской культуры, 2014. С. 310–321.
5. Сахарных Д.М. Развитие удмуртского национального интернета: преодоление барьеров // Журналистика Удмуртии: история и современность. К 100-летию удмуртской национальной периодической печати. Ижевск: УдГУ, 2006. С. 177–182.
6. Сахарных Д.М. Удмуртский национальный интернет: проблемы развития коммуникативной среды // Коммуникативная природа человека (Первые Петраковские чтения). Материалы Российской научно-теоретической конференции 31 января 2006 г., УдГУ. Ижевск: Издательство Удмуртского университета, 2006. С. 62–64.
7. Histoire sociale des langues de France / Sous le direction de G. Kremnitz. [s.p.]: Presses Universitaires de Rennes, 2013.
8. Kymlicka W. Politics in vernacular: Nationalism, Multiculturalism and Citizenship. Oxford: Oxford University Press, 2001.