

Интернет по ту сторону цифр

Сборник статей конференции

Издательские решения
По лицензии Ridero
2018

УДК 31
ББК 60
И73

Авторы: Бельская Е., Бородулина А., Герасимова А., Глазков К.,
Дорофеева О., Карасева А., Королева Д., Ляликова Д., Миронова Т.,
Орехов Б., Паукова А., Пишлëгер К., Томилова Е., Тренихина М.,
Фомичева Т., Чумакова В.

Шрифты предоставлены компанией «ПараТайп»

Интернет по ту сторону цифр : Сборник статей конференции / Е.
И73 Бельская [и д. р.]. – [б. м.] : Издательские решения, 2018. –
238 с. – ISBN 978-5-4490-4731-1

Интернет – это связи, сети, данные, коды, интерфейсы, инфраструктура. И цифры. Большая часть интернета легко переводится в количественные показатели. Но всегда остаётся что-то, что мы не сможем понять, применяя только количественный подход. Признавая важность цифр, мы предлагаем посмотреть, что происходит до того, как реальность оказывается посчитанной и уложенной в круговую диаграмму.

УДК 31
ББК 60

16+

В соответствии с ФЗ от 29.12.2010 №436-ФЗ

ISBN 978-5-4490-4731-1

САМОРЕПРЕЗЕНТАЦИЯ СООБЩЕСТВ, ГОВОРЯЩИХ НА НАЦИОНАЛЬНЫХ ЯЗЫКАХ РОССИИ

Об авторе: Борис Орехов – доцент факультета гуманитарных наук
НИУ ВШЭ, nevmenandr@gmail.com.

В статье рассказывается о том, как интернет-среда задаёт новые рамки для сообществ носителей малых языков России, существующих в ситуации конкуренции с престижным и официальным русским языком. Некоторые специфические для онлайн-площадок метрики (такие, как объём соответствующих разделов Википедии) способствуют оформлению новых сфер конкуренции данных сообществ между собой, что во многом противоречит идеологическим установкам, с которыми создавались эти площадки. Наряду с новыми показателями витальности языка в виде его интернет-презентации мы получаем и новые формы активности сообществ носителей языков, по-своему переформирующих систему координат онлайн-среды.

Ключевые слова: миноритарные языки, Википедия, социальные сети, малые сообщества, социальный капитал.

В нашей статье речь пойдет об одном из ответвлений проекта, который был посвящен исследованию представленности в интернете языков народов России и который на данный момент можно признать завершенным. Языков этих довольно много, и в фокусе нашего внимания оказались не все – но самые крупные из них, с высокой вероятностью встречающиеся в интернете, были проанализированы. В результате исследования для каждого языка были получены разнообразные количественные характеристики: сколько сайтов, сколько сообществ в соци-

альных сетях, коммуникация в которых ведется на том или ином национальном языке, сколько статей в соответствующем языковом разделе Википедии и некоторые другие данные¹.

Основной вывод, который можно сделать из этих количественных замеров, заключается в том, что законы, по которым существует язык в онлайн-среде, не являются простой функцией от законов, обычно действующих на язык как объект изучения социолингвистики². Речь идет прежде всего о таком важнейшем показателе, как число носителей языка. Можно было бы ожидать, что самый распространённый (по этому показателю) язык будет лучше всего представлен и в интернете — однако это не так. Язык с меньшей витальностью может быть шире распространён в сети, чем язык с большим числом носителей.

Так, луговой марийский язык насчитывает более 300 тыс. носителей, а его ближайший сосед — горномарийский, — около 20 тыс. носителей. Несмотря на это, к настоящему моменту число статей в горномарийской Википедии выше³, чем в луговой марийской⁴. Носители горномарийского, очевидно, более активно действуют в интернете (что связано с их большей сплоченностью). Подобная тенденция, вероятно, может проявляться не только в случае интернет-сообществ, но и для других малых коллективов. Однако число говорящих всегда было основным способом количественной оценки всего, что происходит с языком в плоскости социолингвистики.

¹ См.: [<http://web-corpora.net/wsgi3/minorlangs/>].

² Orekhov, B., Krylova, I., Popov, I., Stepanova, E., Zaydelman, L. Russian Minority Languages on the Web: Descriptive Statistics // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог» (Москва, 1—4 июня 2016 г.). Вып. 15 (22). М.: РГГУ, 2016. С. 452—461.

³ См.: [<https://mrj.wikipedia.org/wiki/>], 10240 статей.

⁴ См.: [<https://mhr.wikipedia.org/wiki/>], 9652 статьи.

Особенное внимание мы уделяли таким формам репрезентации малого языка в интернете, как создание и администрирование пабликов во Вконтакте¹, написание статей для Википедии в соответствующем языковом разделе, а также разработка дополнительного программного обеспечения, позволяющего корректно воспроизводить графику языка (программные клавиатуры). Важность последнего пункта обусловлена тем, что большинство малых языков России имеют алфавиты, отличным по составу букв от гражданского кириллического, и кроме того содержащие уникальные символы.

Количественные исследования позволяют отслеживать, как проявляется активность носителей малого языка. Особенно хорошо она видна в Википедии, где в постоянно обновляемой статистике отмечаются новые статьи и правки, а участники сообщества постоянно обсуждают ситуацию. Заметны здесь также и статистические выбросы — резкие наращивания числа статей за счет т.н. «ботозаливок» — автоматизированных операций, увеличивающих статистику размеров того или иного раздела в энциклопедии². В последнее время этот вид активности ха-

¹ Krylova, I., Orekhov, B., Stepanova, E., Zaydelman, L. «Languages of Russia: Using Social Networks to Collect Texts», in Braslavski, P. et al. (eds.) *Information Retrieval. 9th Russian Summer School, RuSSIR 2015, Saint Petersburg, Russia, August 24–28, 2015, Revised Selected Papers*. Volume 573 of the series Communications in Computer and Information Science, p. 179–185. Springer, 2016.

² В архиве башкирской Википедии можно найти обсуждение ботозаливок на русском языке: [<https://goo.gl/gtkE9b>]. Из них следует, что в целом активисты Википедии относятся к этому явлению положительно: «Благодаря ботозаливкам Башвики в свое время выжил и дожил до своего 10 летия. В то время никто серьезно не принимал Башвики и статьи изредка писали 2–3 участника. Ботозаливки привлекли внимание общественности. Тогда и таким статьям были рады» (*орфография и пунктуация авторские*).

рактен именно для малых языков (в данном случае я распространяю это понятие за пределы России, т.е. имею в виду и малые языки других стран, — например, себуанский или вайский) и представляет собой раздувание числа статей за счет автоматического их заполнения; с точки зрения традиционного подхода к оценке объема статей, учитывающего только написанное людьми, это явление следует считать искусственным. Именно поэтому на первых строках рейтинга объема Википедий фигурируют языки, не принадлежащие к числу мировых (например, тот же себуанский).

Примерно та же ситуация наблюдается при изучении языков народов России. В течение последних пяти лет мы видим соревновательный рост числа статей для разделов, привязанных к языкам, титульным для некоторых регионов РФ¹. Речь, прежде всего, идет о национальных республиках (например, тюркоговорящих).

Здесь возникает вопрос, на который количественное исследование ответить не может: каковы причины такого поведения языкового сообщества в интернете? Очевидно, из первичной интенции, сформулированной для Википедии её создателями, эта тенденция выводится не вполне тривиальным образом: теоретически Википедия — это интернет-энциклопедия, призванная накапливать и сохранять знания. Это положение отражено в различных текстах авторства как создателей энциклопедии, так и интернет-активистов, эту идеологию пропагандирующих. Центральный лозунг всей этой концепции — доступность знаний для всех. Соответственно, если некоторый субъект этой системы не говорит на распространённом мировом языке, а владеет

¹ Орехов Б. В., Решетников К. Ю. Государственные языки России в Википедии: к вопросу о сетевой активности миноритарных языковых сообществ // Настройка языка: управление коммуникациями на постсоветском пространстве. М.: Новое литературное обозрение, 2016. С. 263—281.

только малым языком, то интернет-энциклопедия должны быть создана и на его языке, чтобы и он мог ознакомиться с накопленной информацией. Джимми Уэйлс описал Википедию как «усилия по созданию и распространению свободной энциклопедии самого высокого качества для каждого человека на планете на его собственном языке»¹.

В то же время для носителей малых языков России эта идея очевидно бессмысленна, так как почти все они являются по крайней мере билингвами, т.е. умеют говорить и читать по-русски². Поэтому на них не распространяются те ограничения в получении знаний, которые могли бы быть связаны с локальными языковыми возможностями. В связи с этим причин создавать википедийный языковой раздел на таком языке, будь то башкирский, удмуртский или хакасский, строго говоря, нет. Однако подобному прагматическому соображению противостоит культурная составляющая, выводящая язык из чисто утилитарного коммуникативного инструмента в сферу факторов, формирующих идентичность сообщества.

Следует выдвинуть гипотезу, что в наращивании количества статей Википедии за счет их шаблонизированного воспроизведения мы наблюдаем стремление завязать рейтинг соответствующего языкового раздела с тем, чтобы репрезентировать свое сообщество и культуру. Использование для этого авторитетной интернет-площадки, первоначально имевшей другие задачи, предстает при этом не вполне целевым использованием ресурса, оказывающимся, тем не менее, вполне оправданной мерой при формировании статуса своей социальной группы и ее культуры. Сами администраторы языковых разделов видят в такой мере стимулирование активности вокруг их проекта, но конкретный механизм позитивного влия-

¹ Википедия // Википедия. [2017–2017] [<http://ru.wikipedia.org/?oldid=87864983>].

² См. данные Переписи 2010 г.

ния на Википедию ими обычно не раскрывается: «Одним из наиболее эффективных средств для быстрого рывка в развитии раздела является создание ботов (программ, выполняющих автоматически и/или по заданному расписанию создание заготовок статей, рутинные правки, исправление технических ошибок и т.д.) и соответственно применение ботозаливки (массовой распланированной по дням и количеству заливки заготовок статей)»¹. Соревновательный характер подобного раздувания языкового раздела в заявлениях активистов интернет-энциклопедии обычно не проговаривается открыто. В то же время акцент на количественных показателях (сколько статей создано, на каком месте в перечне разделов по числу статей находится данный языковой раздел) обязательно в них присутствует и очевидно осознается как центральный и в концептуальном, и в идеологическом плане. Случайный пример из многих подобных: «К сведению, интернет-энциклопедия на башкирском языке входит в первую сотню среди 288 языковых разделов Википедии и занимает 82 место по общему количеству статей»².

За событиями, таким образом, стоит не подразумевавшаяся изначально функциональность Википедии, связанная не столько с накоплением или распространением знания, сколько с выстраиванием виртуальной культурной иерархии, заметное место в которой занимают малые языки и культуры; однако занимают они его не в силу исторической обусловленности или значимости в современном мире, а в результате фетишизации количественных показателей. «Виртуальным» этот рейтинг был назван мной потому, что он существует исключительно в сознании интернет-активистов исследуемых сообществ.

¹ См.: [http://lnka.ru/glav_lenta/9864-dobrozhelatel'naya-lezginiskaya-vikipediya.html].

² См.: [http://kulturaufa.ru/odbor/news/?ELEMENT_ID=5281].

Аналогичным образом любопытные вещи происходят и в среде Вконтакте¹. Оговоримся сразу, что наши наблюдения носят предварительный характер, и прежде чем на них можно будет всерьез опираться, следует провести типологическое исследование аналогичных процессов в русской части социальной сети, где гипотетически могла бы возникать конкуренция между локальными сообществами, говорящими на одном языке – например, жителями соседних городов или регионов. Представляется, тем не менее, что многие ресурсы просто дублируют друг друга, так что тематика некоторых пабликов практически идентична. Этот процесс изоморфен тому, что мы наблюдаем в Википедии, хотя происходящее во Вконтакте, формализовать намного труднее².

Итак, что касается представления малых языков в интернете, то в тех случаях, когда речь идёт об авторитетных площадках вроде Википедии, среди представителей сообщества носителей того или иного языка наблюдаются активности, прямо не связанные с декларированной функциональностью этих площадок. Важнейшим показателем оказывается количественная измеримость этих активностей и возможность повысить позиции своего языка в рейтингах, учитывающих эти измерения.

Представляется, что конкуренция в Сети не противоречит конкуренции в офлайн-среде, а дополняет и расширяет ее, вводя новые метрики и поля для взаимодействия, а также открывая новые возможности для построения иерархий и инициации соревновательности.

Библиография

¹ *Orekhov, B., Krylova, I., Popov, I., Stepanova, E., Zaydelman, L.* Russian Minority Languages on the Web: Descriptive Statistics.

² См. данные нашего исследования на [<http://web-corpora.net/wsgi3/minorlangs/>].

Орехов Б. В., Решетников К. Ю. Государственные языки России в Википедии: к вопросу о сетевой активности миноритарных языковых сообществ // Настройка языка: управление коммуникациями на постсоветском пространстве. М.: Новое литературное обозрение, 2016. С. 263–281.

Krylova, I., Orekhov, B., Stepanova, E., Zaydelman, L. «Languages of Russia: Using Social Networks to Collect Texts», in Braslavski, P. et al. (eds.) *Information Retrieval. 9th Russian Summer School, RuSSIR 2015, Saint Petersburg, Russia, August 24–28, 2015, Revised Selected Papers*. Volume 573 of the series Communications in Computer and Information Science, p. 179–185. Springer, 2016.

Orekhov, B., Krylova, I., Popov, I., Stepanova, E., Zaydelman, L. Russian Minority Languages on the Web: Descriptive Statistics // Компьютерная лингвистика и интеллектуальные технологии: По материалам ежегодной Международной конференции «Диалог» (Москва, 1–4 июня 2016 г.). Вып. 15 (22). М.: РГГУ, 2016. С. 452–461.