

РУССКИЙ язык И НОВЫЕ ТЕХНОЛОГИИ

Сост. Г.Ч. Гусейнов

Под ред. М.В. Ахметовой, В.И. Беликова

Москва
Новое литературное обозрение
2014

ВВЕДЕНИЕ

РУССКИЙ ЯЗЫК И НОВЫЕ ТЕХНОЛОГИИ: СОВРЕМЕННОЕ СОСТОЯНИЕ И ПЕРСПЕКТИВЫ ИССЛЕДОВАНИЙ

Масштабная информатизация общества и изменения в геополитическом статусе русского языка пришлись на один исторический период. Конец 1980-х и начало 1990-х гг. стали временем, когда два поначалу независимых друг от друга процесса начали формировать современную социокультурную и социолингвистическую повестку в самом широком смысле этих слов.

Распад СССР перекроил карту Евразии, по-новому оформив границы ареалов распространенности русского языка, одновременно сделав его носителей «национальным меньшинством» (не всегда желанным) в таких регионах, как Туркменистан, Украина, Узбекистан и др. В 1990-х гг. русский язык, оставшись *lingua franca* для стран СНГ и большей части Восточной Европы, по сути, еще более укрепил позиции в качестве языка международного общения. Случившееся напоминает ситуацию с дополнительной эмиссией акций на фондовом рынке, когда размер активов компании и их цена не изменяются, но количество долей в этих активах растет. Таким же образом примерно на прежнем уровне осталось число носителей русского языка, но количество государств, для которых русский язык стал необходимым коммуникационным ресурсом, возросло на порядок. Были пробуждены к жизни множественные сложные социальные и, как следствие, социолингвистические процессы. Язык русских Казахстана (и некоторых других бывших союзных республик) стал языком диаспоры, то есть относительно изолированным, самостоятельно развивающимся социолектом (в нем уже появились специальные обороты, вроде *маякнуть кому-либо* в значении «позвонить, сделать дозвон»), а бывшие до конца 1980-х исчезающе слабыми связи метрополии с русской эмиграцией интенсифицировались, что, конечно, тоже имело свою лингвистическую проекцию для обеих сторон. Так, для чуткого уха раздражающими стали обороты из русита (диалекта русской диаспоры Израиля) вроде *из этого мы учим* (в значении «отсюда мы узнаем») или *различать между* («различать что-то»)¹.

В самой России также произошло достаточно внушительных событий, чтобы облик русского языка 1990-х и 2000-х гг. приобрел ин-

¹ Автор благодарит З. А. Махмутова и А. Л. Поляна за помощь в подборе иллюстративного материала.

дивидуализированные черты. Серьезная перетряска структуры общества актуализировала в национальном языке лексические пласты из сферы общения прежде маргинальных социальных групп. Особенное влияние на язык оказала криминализация общества в 1990-х гг., легитимизировавшая в активном словаре русскоязычных людей *наезд*, *стрелку* и целые ряды подобных лексем, что даже не всеми оценивалось негативно. Так, много страниц посвятивший размышлениям на эту тему писатель В. Пелевин говорил, что «в лексике братков есть огромная сила, и русский язык, захиревший в речи интеллигентов, воскрес в блатном базаре, возродившем первозданность понятий жизни и смерти».

Появились и многочисленные слова, называющие новопришедшие в жизнь реалии (*зачистка*, *распил*), язык столкнулся с необходимостью осваивать заимствования, до сих пор появлявшиеся в таком количестве, кажется, только при Петре I.

Параллельно с этим дешевели и приходили в сферу частного пользования компьютеры, а вслед за ними — доступ к всемирной компьютерной сети Интернет. Сравнительно долгое по меркам новейшей истории время Сеть оставалась сферой приложения усилий специалистов (программистов и иных представителей компьютерных профессий), что было связано и с тем, что благосостояние рядовых российских граждан на тот момент не всегда позволяло иметь дома соответствующую технику, и с тем, что на ее освоение от старшего поколения требовались усилия, сравнимые с теми, которые нужно потратить на изучение нового языка. На заре развития глобальных компьютерных сетей одним из пионеров Рунета Романом Лейбовым была придумана виртуальная личность пенсионера по имени Май Иванович Мухин, освоившего Интернет и создавшего персональную страничку². Это казалось настолько необычным, что за персонажем закрепилась аббревиатура ПППвППП (Первый и Последний Пенсионер в Повсеместно Протянутой Паутине), а сам пенсионер удостоился похвалы в публичном выступлении президента Эстонии Леннарта Мери [Скородумова 2004]. Прошли годы, прежде чем стало ясно, что тесные взаимоотношения с компьютерным интерфейсом — это неизбежность для каждого, за это время молодые люди активно осваивали компьютер и успели сформировать облик (в том числе и языковой) молодого Рунета.

Если распад СССР главным образом провоцировал дезинтеграцию на разных уровнях, то Интернет стал своего рода компенсаторным фактором. Сначала электронная почта и чаты, затем форумы, блоги и, наконец, социальные сети обеспечивали удобство коммуникации между разобщенными в силу исторических причин людьми.

²http://www.cs.ut.ee/~roman_l/muxin_old.html.

Роскошь человеческого общения стала доступнее. Государственные границы и расстояния перестали быть помехой иллюкутивности. С момента своего появления интернет-среда накапливала массу, ставшую критической в середине 2000-х, когда были узаконены или порождены новые формы и средства коммуникации (эмотиконы, или «смайлики», ответы вида «+1» и пр.), специфические формы языковой игры (*превед, кросавчег!*). Что-то из этого было оперативно осмыслено лингвистами и превратилось в академическую рутину, а что-то нуждается в самом серьезном анализе до сих пор. Так, не вполне ясен статус речи, которую можно наблюдать в интернет-мессенджерах, чатах и прочих средствах моментальной доставки сообщений. С одной стороны, формально она является письменной: использует графические знаки, притом часто расширяя традиционный смысловой репертуар символов. С другой стороны, такая речь во многом спонтанна и отвечает многим тенденциям устного модуса.

Но коммуникацией дело не ограничилось. Интернет стал гигантским хранилищем текстовой информации, легкодоступной и востребованной, таким образом обогатив актуальное референтное поле русской культуры, введя в него не только привычную школьную классику, но и разножанровые документы, не самые распространенные ранее языковые пласты и речевые произведения. В потенциальном доступе у носителя языка и в фокусе внимания ученого-лингвиста оказались колоссальные объемы, которые поменяли и угол зрения на тексты на данном языке, и аргументы, которыми оперируют ученые. Так, пользователь Интернета способен легко отыскать попавшуюся ему на глаза цитату и ознакомиться с первоисточником, обогатив таким образом свой культурный фон. Легкость, с которой любой человек получает такую возможность, без сомнения, провоцирует ею пользоваться. Результаты этого видны уже по тому числу энтузиастов-любителей (иногда агрессивных), которые создают свои сообщества вокруг интересных им областей (скажем, русских культурных древностей) и пропагандируют свои не вполне логичные и последовательные «теории», опираясь при этом на видимость фактов и труды авторитетов. Такие упражнения все сложнее становится отличать от действительно научных трудов по формальным признакам, в которых ранее отражалась кропотливая работа ученого с источниками. В современных условиях с источниками зачастую может справиться и непрофессионал.

Ученому-лингвисту электронные коллекции текстов во многом вернули предмет его научных занятий (см. об этом: [Плунгян 2008]), обратив от полунаучных абстракций к конкретному речевому материалу, попутно снабдив возможностью применять к нему количественные методики точных и естественных наук. В новую эпоху стало возможно с высокой степенью достоверности судить о том, что встре-

чается в языке чаще, что реже и что за этим стоят системные причины, а не простая случайность.

Объять необъятные текстовые пространства помогают поисковые машины, которые вообще серьезно видоизменили структуру информационного поля и, в частности, сформировали новый взгляд на язык и на речевую культуру. Показательной стала распространенная привычка проверять сложные случаи правописания с помощью тех же поисковых систем. Как правильно, «предвидеть» или «предвидить»? В 2000-х гг. ответ на этот вопрос давали Яндекс и Google: тот вариант, который встречался в Интернете большее число раз, признавался грамотным. Такая вера в коллективный разум может показаться наивной, ведь ничто не мешает активным пользователям Интернета чаще ошибаться в написании хотя бы некоторых слов. Но на самом деле за ней стоит вполне разумное понимание языка как функции от совокупности текстов. Если так, как пишут по-русски, неправильно, то что же такое «правильно»?

Сейчас, рассматривая конкретные языковые изменения, зачастую трудно отделить следствия одного из этих глобальных процессов от следствий другого. Информатизация и геополитические катаклизмы исторически тесно переплетены. Но несомненно то, что русский язык в последние десятилетия меняется. Язык (любой язык) меняется постоянно под действием внутренних (не вполне изученных) и внешних факторов. Обратное — скорее исключение, редчайшие случаи географической или социальной изоляции небольших коллективов. Однако в последние десятилетия русский язык развивается особенно интенсивно, включаясь в общемировые процессы и участвуя в них на равных (или почти на равных) с другими языками международного общения. Очевидно, что именно либерализация советского общества позволила русскому языку избежать интернет-изоляции и начать играть ту роль, которую он сейчас играет в Сети и в офлайне, — роль глобального ресурса.

Коснемся подробнее одного примечательного факта. 19 марта 2013 г. сайт W3Techs, занимающийся анализом веб-сайтов, сообщил о том, что русский язык вышел на второе место среди языков Интернета³, уступая только английскому. Эта новость имеет прежде всего два аспекта: технический и социолингвистический.

Чтобы выяснить, на скольких страницах в Сети употреблен тот или иной язык (скажем, русский), нужно загрузить и просмотреть каждую такую страницу. «Смотреть», конечно, будет не человек, потому что, хотя Интернет и создан для людей и их удобства, своими объемами он очень быстро перерос человеческие возможно-

³ http://w3techs.com/blog/entry/russian_is_now_the_second_most_used_language_on_the_web.

сти. На 2009 г. объем Сети составлял около 500 экзабайтов, то есть 500 млрд гигабайтов, и с тех пор до того момента, когда W3Techs осуществил свое исследование, увеличился не менее чем в 1,5 раза. Чтобы собирать информацию с веб-сайтов, существуют специальные программы-роботы, которые называются слайдерами (то есть «пауками», бегающими по Сети) или краулерами. Они скачивают веб-документы и передают их для обработки и анализа другим программам. Но даже программы, запущенные на обычном компьютере, будут делать такую работу слишком долго. Нужна внушительная техническая база, парк объединенных под общую задачу серверов; нужно решать и сопутствующие логистические задачи, чтобы обеспечить этим серверам бесперебойную работу. С этими сложностями сталкиваются в первую очередь поисковые системы вроде Яндекса и Google, которые сканируют Интернет и предоставляют информацию о его содержимом по запросам пользователей. Если не обладать возможностями этих крупных компаний, то качественный «обход» Сети вряд ли получится. Чтобы просеять Веб за конечное время, поневоле придется давать сайтам приблизительные оценки. Об этом пишут и сами сотрудники W3Techs в разделе часто задаваемых вопросов⁴, указывая, что во многих случаях ограничиваются только главной страницей известного им сайта и не заглядывают глубже. Соответственно, как распределены языки на таких сайтах, если сайт многоязычный, можно судить лишь в самых общих чертах, что, конечно, влияет и на точность выводов исследования.

Кроме того, известно, что в последнее время основная жизнь в Интернете переместилась в социальные сети и микроблоги. Каждую секунду в Facebook, «ВКонтакте», Twitter создаются и транслируются новые записи. Однако даже крупные поисковые системы испытывают трудности с тем, чтобы своевременно и со всей полнотой фиксировать эти действия пользователей: социальные сети технически специально устроены, чтобы самостоятельно распоряжаться своим контентом, не делясь этим лакомым куском с конкурентами. Надежды на то, что W3Techs научились сканировать Facebook лучше, чем Google или Яндекс, невелики, а ведь именно внутри социальных сетей потенциал интернет-жизни любого языка раскрывается в полной мере.

Есть и другие технические нюансы. Например, когда мы пытаемся выявить, сколько в Интернете страниц, содержимое которых написано на том или другом языке, должны ли мы (как это делают те же поисковики) «склеивать» дублирующие друг друга страницы и отображать сайты, созданные только для того, чтобы обмануть поискового робота? Ответов на эти вопросы на сайте W3Techs обнаружить

⁴ <http://w3techs.com/faq>.

не удалось. Нужно ли говорить, что и это серьезно влияет на конечный результат подсчетов.

Наконец, не стоит сбрасывать со счетов и такое обстоятельство, как трудность автоматического определения языка. В тот момент, когда краулер скачивает страницу, он передает ее для дальнейшей обработки распознающей язык программе, и гарантии, что та определит язык безупречно, быть не может. Разумеется, сложностей в том, чтобы отделить японский язык от французского, не возникнет: у этих языков разная письменность. Но сумеет ли машина верно увидеть, русский язык перед ней или болгарский, уже большой вопрос. Так, в 2006 г. компания «СУП», желавшая взять под свой контроль русскоязычную часть «Живого журнала», даже не стала решать проблему вычленения блогов по языковому принципу и просто начала обслуживать весь кириллический сегмент блог-платформы, включая блоги на украинском, белорусском, сербском и том же болгарском. Системы распознавания вообще чрезвычайно сложны и трудоемки в разработке, с ними могут не вполне справляться даже крупные компании. Тот же Google предлагает при поиске показывать страницы «только на русском» и часто ошибается.

Одним словом, к подсчетам W3Techs нужно относиться с большой осторожностью, особенно учитывая, что, по их данным, русский язык во второй половине марта 2013 г. вышел на второе место, незаметно обойдя не какой-то другой язык, а немецкий. W3Techs же является одним из направлений деятельности именно немецкой консалтинговой компании Q-Success. Разумеется, не следует подозревать в этом подтасовку данных. Просто вполне вероятно, что немецкая компания лучше всего осведомлена о немецкоязычной части Интернета и исследует ее с особенной тщательностью. Однако проверить выводы W3Techs независимому эксперту будет сложно, так как слишком велики сопутствующие инфраструктурные издержки.

Но даже если принять во внимание все эти трудности, одному из результатов исследования W3Techs верить можно: русский язык быстрыми темпами наращивает свое присутствие в Интернете. Похожая проблема — проблема популярности — является предметом заочного спора между двумя компаниями — разработчиками браузеров: Google, агрессивно продвигающей на рынке свой продукт Chrome, и Microsoft, создателем Internet Explorer, позиции которого среди программ для навигации по Интернету долгое время были непоколебимыми. Две компании, занимающиеся интернет-аналитикой, — statCounter и Net Applications — имеют радикально разные мнения о том, какой браузер сейчас востребован у большего числа пользователей. Первая отдает предпочтение Chrome, а вторая — Internet Explorer, предоставляя и Google, и Microsoft право называться создателями самого популярного в мире браузера. Однако и statCounter,

и Net Applications фиксируют ускорение роста пользовательской базы Chrome и замедление в росте интереса к Internet Explorer. Так что независимо от текущего состояния дел победа Chrome в этой гонке всего лишь дело времени.

Так что важнее не сколько десятых процента сейчас русский язык отвоевывает у своих «конкурентов», а какими темпами его доля растет в Интернете. И, по данным W3Techs, этот показатель для русского языка выглядит оптимистично⁵.

Однако устойчивый рост русский язык обнаруживает пока только в Интернете. По распространенности в мире он занимает 8-ю позицию с числом говорящих около 160 млн человек. Из этого следует, что все больше носителей русского языка знакомятся с Интернетом, а большинство людей, которые говорят на китайском, испанском, арабском и других языках, обходящих русский по численности, пока еще далеки от высоких технологий (или их интернет-активность в недостаточной степени учитывается W3Techs). На этом примере мы в чистом виде наблюдаем так называемый эффект низкого старта: уровень «информатизации» постсоветского общества (основного поставщика русскоязычного контента) изначально был очень низким, гораздо ниже, чем у немецко-, испано- и франко-, а уж тем более англоязычного социумов. В результате все остальные языки уже, скажем так, выбрали свой ресурс наполнения Сети контентом, а русскоязычные пользователи, по сути, только начинают поставлять в Интернет свои тексты. Об этом свидетельствует и уже упоминавшийся главный показатель распространенности: динамика роста. Если у немецкого языка она замедляется, то просто потому, что большинство немецкоговорящих уже имеют доступ к Сети и расти этой аудитории некуда. Когда-нибудь наступит такое время, когда каждый житель планеты будет с обязательностью одновременно и пользователем Интернета, тогда выровняется число говорящих и представленность языка в Вебе, то есть русский много уступит китайскому, на котором говорит более миллиарда человек, притом что китайский в рейтинге W3Techs занимает только 6-ю строчку. Возможно, что доля китайского в Интернете и сейчас выше. Если это не так, то дела с информатизацией в Китае обстоят гораздо хуже, чем в России. Впрочем, об активном воздействии в Китае государства на происходящие в Интернете процессы хорошо известно, возможно, что оно серьезно сдерживает рост национального сегмента Сети. Не случайно агонизирующие режимы в Тунисе и Египте пытались перекрыть доступ к Интернету в обстановке набирающей обороты революции.

Как бы там ни было, доступ к современным технологиям — это безусловное достижение прогресса в самом лучшем смысле этого

⁵ http://w3techs.com/technologies/history_overview/content_language/ms/y.

слова, а та положительная динамика, которую демонстрирует присутствие русского языка в Интернете, прежде всего свидетельствует о растущем благосостоянии российского общества.

Поэтому рождение своих коммуникативных привычек у пользователей Сети было неизбежным. Но Интернет не только заставляет язык меняться. Он еще и помогает изучать лингвистические аспекты происходящих изменений. В Сети хранится наиболее обширная и динамически пополняемая электронная коллекция текстов на русском языке, в частности лучше, чем где бы то ни было, в Интернете представлены тексты последних лет. Многие из этих речевых произведений датированы, какие-то имеют географическую привязку, мы можем также почерпнуть информацию о поле и возрасте автора, о других социолингвистически значимых параметрах. Такой материал, конечно, открывает (и уже открыл) богатейшие возможности для установления языковых закономерностей, подкрепленных статистически релевантными объемами. Именно сейчас человек выходит на новый уровень в понимании если не того, как мы говорим, то в понимании того, как к этой проблеме подступиться.

По сути, проблема русского языка и Интернета в исследовательской плоскости распадается на три автономных аспекта. Во-первых, это русский язык Интернета, те самые новые формы и средства коммуникации, которые были вызваны к жизни созданными в последние десятилетия каналами передачи информации. Новые медиа переконфигурировали сообщество говорящих, изменили иерархию участников общения, привели к появлению специальной этики коммуникации и речевых сигналов особой семантики. Расхожим стало оправдывающее многое выражение «это Интернет, детка, здесь могут и послать». Стоящая за таким положением вещей идеология и сопряженные с ней практики письма, на которые накладывается особый визуальный ряд, по-прежнему требуют осмысления. Ведь даже чтобы сказать, что в последние годы мы столкнулись с невиданными в истории человечества процессами, необходимо предварительно провести работу по поиску более или менее отчетливых исторических параллелей. Такая работа идет и остается актуальной⁶, так как мы всё еще далеки от полноты понимания того, что произошло (и происходит вокруг нас) в новшествах новых медиа, в механизме распространения мемов и различных способов языковой игры.

Во-вторых, это русский язык в Интернете. Помимо всех тех новшеств, которые были привнесены в коммуникацию новыми технологиями, доступные в Сети оцифрованные тексты в первую

⁶ См., в частности, сборник норвежского научного проекта The Future of Russian «From Poets to Padonki. Linguistic Authority and Norm Negotiation in Modern Russian Culture» (Bergen, 2009).

очередь — это максимально подробный портрет языковой системы в целом, или, как говорят лингвисты, «корпус», лингвистическая справочная система, способная послужить надежным базисом для языковедческих исследований. Использование Сети как корпуса на Западе уже превратилось в специфическую исследовательскую дисциплину *web as corpus* (см., например: [Hundt et al. 2007; Gatto 2009]), но в России пока остается недооцененным.

Наконец, совершенно неисследованную, незаслуженно забытую область разысканий представляют собой языки национальных меньшинств России, их представленность в Интернете и конкуренция с доминантным языком региона — русским. До самого последнего времени вопросы о взаимодействии русского языка и миноритарных языков России не были даже поставлены, хотя понятно, что в современных условиях именно сетевая жизнь во многом определяет уровень витальности языка и его перспективы в условиях постиндустриального общества.

Обсуждение методологии этих трех аспектов взаимоотношения языка и Интернета (и имманентно, и в связи с сопутствующими вопросами) на сегодняшний день составляет наиболее актуальную задачу на том научном поле, в котором взаимодействуют лингвистика, социология и культурология, пытающиеся совместными усилиями достичь понимания природы происходящего в Интернете. Если первый из названных аспектов уже имеет некоторую индивидуальную историю, второй пока робко, но все же набирает обороты, третий все еще в большей степени перспективен, нежели насущен. Но все три в ближайшем будущем составят мейнстрим научного изучения Интернета с лингвистической точки зрения.

Литература

Плунгян 2008 — *Плунгян В. А.* Корпус как инструмент и как идеология: о некоторых уроках современной корпусной лингвистики // Русский язык в научном освещении. 2008. № 16 (2). С. 7-20.

Скородумова 2004 — *Скородумова О. Б.* Виртуальная личность и свобода (к проблеме социокультурных истоков понимания свободы в Интернете) // Вестник Московского университета. Сер. 7. Философия. 2004. №2. С. 75-96.

Gatto 2009 — *Gatto M.* From Body to Web: An Introduction to the Web as Corpus. Bari: Editori Laterza, 2009.

Hundt et al. 2007 — *Corpus Linguistics and the Web / Ed. by M. Hundt, N. Nesselhauf and C. Biewer.* Amsterdam; New York: Rodopi, 2007.