

*В. А. Плунгян*

## **ЗАЧЕМ НУЖЕН НАЦИОНАЛЬНЫЙ КОРПУС РУССКОГО ЯЗЫКА? НЕФОРМАЛЬНОЕ ВВЕДЕНИЕ<sup>1</sup>**

Русский язык, его отличие от других языков, его прошлое и настоящее, его судьба на рубеже XX и XXI столетий, его будущее, возможность радикальных изменений в его строе и в сфере употребления — все эти темы практически необозримы и вызывают интерес не только у профессиональных лингвистов. Тем не менее, если вопросы по этому поводу задавать именно лингвистам, то с большой вероятностью в их ответах сразу прозвучит слово «корпус».

Действительно, в науке о языке подобные фундаментальные проблемы с недавнего времени обсуждаются в основном с помощью такого инструмента, как *корпус языка*. Это очень эффективный и полезный инструмент (которым могут пользоваться далеко не только узкие специалисты), особенно в том случае, когда корпус является большим по объему и полным по охвату материала, т. е. представляет собой так называемый Национальный корпус языка.

Теперь следует пояснить, что такое корпус вообще и Национальный корпус в частности. Корпус некоторого языка — это, в первом приближении, собрание текстов на данном языке, представленное в электронной форме и снабженное научным аппаратом. Аппарат, «встроенный» в корпус, обычно называется «разметкой», или «аннотацией», корпуса; корпус тем лучше, чем полнее и совершеннее его аннотация. Собственно, наука о корпусах (а таковая давно уже существует и называется «корпусная лингвистика») — это прежде всего наука о том, как сделать хорошую разметку корпуса.

Хорошая разметка, в частности, позволяет быстро и эффективно найти в корпусе те слова, формы и конструкции, которые нуж-

---

<sup>1</sup> Сокращенный вариант данной статьи был опубликован в журнале «Отечественные записки» (2005, № 2: Общество в зеркале языка).

### **Зачем нужен Национальный корпус русского языка?**

ны исследователю. Для этого программа поиска должна понимать как минимум то, какие формы в тексте относятся к одному и тому же слову (например, *ржи* и *рожью*, но не *ржал* и *Ржев*), то есть хотя бы частично «понимать» грамматическую структуру данного языка.

Тем более это понимание необходимо, если мы хотим искать не слова, а формы. Представьте себе такую задачу: найти в достаточно длинном тексте все формы дательного падежа единственного числа. Никакой текстовый редактор с этой задачей, конечно, не справится. Для того чтобы грамматические формы можно было автоматически найти в тексте, этот текст как раз и необходимо предварительно разметить. Иначе поиск возможен только вручную, а это процедура долгая и трудоемкая.

Разумеется, разметка может быть нужна и для решения многих других задач. Не будет преувеличением сказать, что хорошо размеченный текст для специалиста оказывается просто бесценным. Ведь в своей исследовательской работе лингвисты зависят прежде всего от количества и качества собранного материала. У многих еще свежи в памяти те времена, когда примеры выписывались из текста и заносились на карточки. Сейчас карточки ушли в прошлое, но сама процедура выбора примеров из текста делается человеком и с трудом поддается автоматизации. Размеченные корпуса — первый серьезный инструмент, позволяющий существенно ускорить и упростить эту процедуру. Иными словами, то, на что у исследователей предшествующих поколений уходило недели, а то и месяцы напряженного труда, с помощью корпуса можно сделать за минуты (но, разумеется, вначале надо сделать корпус).

Итак, корпус — это электронное собрание текстов, размеченное таким образом, чтобы в нем можно было быстро найти слова и конструкции с заданными грамматическими и другими интересными лингвисту свойствами. Что же такое Национальный корпус?

В данном случае слово «национальный» — своего рода термин, отражающий скорее семантику английского слова «national», чем русского слова «нация». Впервые это определение появилось в названии Британского национального корпуса (British National Corpus, BNC), созданного в 1990-е годы в Великобритании специалистами-лексикографами; это не самый первый электронный корпус, созданный в мире, но один из лучших, крупнейших и наибо-

лее известных. Для британцев слово «национальный» означало в первую очередь «характеризующий британский национальный вариант английского языка» (в отличие от американского, австралийского и т. п.), но, поскольку этот корпус очень быстро стал практически эталоном корпуса вообще, то значение слова «национальный» незаметно изменилось. Национальным корпусом стали называть просто самый большой и представительный корпус, характеризующий язык данной страны в целом. Таким, например, оказался Чешский национальный корпус (Český národní korpus), содержащий практически исчерпывающую коллекцию текстов на чешском литературном языке разных периодов.

Понятно, что Национальный корпус должен быть прежде всего большим: его объем измеряется сотнями миллионов словоупотреблений (для сравнения можно сказать, что, например, полное собрание сочинений Ф. М. Достоевского насчитывает «всего» около 2 миллионов слов). Но, кроме того, он — и это даже важнее — должен быть представительным. Иначе говоря, он должен содержать все типы текстов, представленные в данном языке в данный исторический период, и при этом содержать их в правильной пропорции.

Именно поэтому Национальный корпус русского языка не ограничивается, например, только произведениями художественной литературы XIX и XX века, сколь бы важны они ни были для изучения русского языка. Он содержит и газетные и журнальные статьи разной тематики (от общественно-политических до, например, спортивных), и специальные тексты (научные, научно-популярные и учебные по разным отраслям знания), и рекламу, и частную переписку, и дневники. Словом, в Корпус попадают образцы практически любого существующего в русском языке письменного дискурса — от статьи современного музыкального критика до инструкции по уходу за кактусами, от рассказов Пелевина до справочника по физике. Более того, составители Корпуса хорошо понимали, что для полного и адекватного представления о том, что происходит в современном русском языке (или, если угодно, с современным русским языком) необходимо еще в большей степени расширить рамки Корпуса и включить в него, наряду с письменными текстами, также и записи устной речи. В настоящее время эта работа делается.

Почему образцы устной речи так важно иметь в Корпусе? Люди пишут не так, как говорят; в особенности это различие ощутимо

### **Зачем нужен Национальный корпус русского языка?**

для языков с давней письменной традицией, за время существования которой нормы письменной и устной речи успевают разойтись достаточно сильно. Письменная речь всегда более консервативна; к тому же, например, в русском языке письменная традиция формировалась под сильным иноязычным влиянием (первоначально старославянским, впоследствии немецким и французским), которое гораздо меньше затронуло устную речь. В этом отношении русский язык не исключение: сходным образом дело обстоит во многих европейских письменных языках, от чешского до французского, а также во многих письменных языках Азии с древней литературной традицией.

Таким образом, если мы хотим выявить наиболее динамичные структуры живого русского языка — и если мы хотим хотя бы отчасти заглянуть в будущее русского языка, мы должны обратиться к стихии устной речи, не скованной традицией и нормой. Многие в устной речи поражает — но, с другой стороны, многие конструкции, существующие в современной устной стихии, неожиданно всплывают в документах времен Алексея Михайловича и даже в новгородских берестяных грамотах XII-XIV вв.

Источником устных текстов для нас являются, например, записи радиопередач, интервью, круглых столов и т. п., а также тексты, полученные в более неформальной обстановке, — например, беседы социологов или маркетологов с целевыми группами респондентов на определенную тему (качества какого-либо товара, социальные проблемы, личные обстоятельства участников беседы, и т. п.): в таких ситуациях люди, как правило, говорят достаточно свободно и естественно. Но, не ограничиваясь этим, мы ведем также записи бытовых разговоров (диалогов и монологов) на улице, в учреждениях, в домашней обстановке. В таких записях участвуют группы добровольцев из разных регионов России, так как в Корпусе, конечно, должны быть образцы речи не только жителей крупных столичных городов. Сразу следует добавить, что отдельной проблемой является включение в Корпус образцов не общерусского языка (пусть и в его разговорном варианте), а настоящей диалектной речи. Такие образцы мы тоже надеемся со временем в Корпус включить.

Здесь нельзя не сказать еще об одном типе текстов, активно используемых в Корпусе уже на данном этапе работы. Это тексты так называемой «электронной коммуникации»: переписка по элек-

тронной почте, а также всевозможные чаты, форумы, общедоступные электронные дневники (так называемые «блоги»<sup>1</sup>) и т. п. В такого рода текстах мы имеем дело фактически со спонтанным письменным самовыражением, дающим очень своеобразный гибрид устной и письменной речи. Конечно, даже в форумах и блогах люди продолжают писать не так (или не совсем так), как говорят, но дистанция между письменной и устной речью здесь во много раз меньше. Этим электронная коммуникация и привлекательна для лингвистов: какие-то языковые явления встречаются только или преимущественно в данных текстах. Впрочем, нельзя не заметить, что тексты электронной коммуникации наиболее сложны для обработки и разметки, потому что пишущие в сетевой среде чувствуют себя более свободно не только в отношении языка, но в отношении норм общепринятой орфографии. Орфографическая вариативность электронных текстов возрастает на порядок, в них широким потоком вливаются разговорные дублеты, произносимые на месте литературных, но обычно не отражаемые на письме — такие, как *ща* (и *щас*), *сёдни*, *чё*, *тока*, *ваще*, *кто-нить* — и многое другое (русистам давно и хорошо известное). Интересно, что пишущие часто прибегают к искажению орфографии сознательно, в целях языковой игры. Более того, кажется, что массовые игры с орфографией (которые заслуживали бы отдельного обстоятельного рассказа) как раз и возникли именно в электронной среде как специфическое средство выразительности — если угодно, художественный прием. Существует даже специальная мода, вышедшая из нескольких сетевых сообществ и как пожар распространившаяся по всему русскому Интернету (например, орфографические клише типа *пеши исчо* или *аффтар*) — своеобразный код, по которому посвященные узнают друг друга. Но, конечно, привлекают электронные тексты далеко не только своей орфографией и раскованным синтаксисом, но и, например, лексикой: опять-таки, это (если не считать некоторых молодежных периодических изданий или произведений массовой литературы) один из самых надежных источников для изучения современного сленга — как по количеству

---

<sup>1</sup> Самым известным русскоязычным блогом, является, по-видимому, так называемый «Живой журнал» ([www.livejournal.com](http://www.livejournal.com)), созданный первоначально в США, но ставший весьма популярным в России, в частности, среди гуманитариев и людей «свободных профессий»; впрочем, аналогичных ресурсов имеется довольно много.

### **Зачем нужен Национальный корпус русского языка?**

примеров, так и по обилию максимально естественных контекстов. А сленг, напомним, очень важен для лингвистов, это своего рода «лаборатория языка», где намечаются пути его будущего развития.

Если посмотреть на пропорции разных типов текстов в Корпусе, то окажется, что доля собственно художественных текстов (проза и драматургия<sup>2</sup>) не так уж велика: она составляет около 40%, т. е. меньше половины. При этом в число художественных текстов включается и мемуарная проза, которая очень интересна и показательно с точки зрения языка, но многими исследователями справедливо квалифицируется как жанр, промежуточный между художественной литературой и чистым «non-fiction». Это во многом противоречит установкам традиционной отечественной филологии (вообще явно или неявно ставившей знак равенства между изучением русского языка и изучением русского *литературного* языка, т. е. языка художественной литературы). Сам термин «литературный» язык (обозначающий на самом деле скорее не литературный в прямом смысле, а просто общенациональный нормативный, т. е. *стандартный* язык) отчасти предполагает, что наиболее престижная и «правильная» часть текстов, создаваемых на данном языке — это тексты художественной прозы. Между тем, такая точка зрения как минимум не очевидна и требует доказательств. Верно ли, что язык писателей (в том числе выдающихся писателей) и общенациональный язык — это одно и то же? В общем случае это, конечно же, не так — и тем более это утверждение не соответствует действительности во второй половине XX и начале нынешнего века, когда социальное значение художественной литературы (и, как следствие, идейно-стилистические установки современных постмодернистов) радикально изменились в сравнении с концом XIX — началом XX века. Поэтому в Национальном корпусе русского языка художественные тексты представлены в большом количестве, но они не доминируют. Интересно, что в корпусах европейских языков эта доля еще более низкая и, как правило, не превышает 20%, так что Национальный корпус русского языка всё равно остается одним из самых «литературоцентричных». Современные пи-

---

<sup>2</sup> К сожалению, русские поэтические тексты пока в общем корпусе не представлены — их включение и обработка требуют особых усилий и решения целого ряда сложных проблем, в том числе и технических. В будущем создание корпуса русских поэтических текстов, конечно, планируется.

сатели — во всяком случае те, что считаются наиболее интересными — скорее выступают как экспериментаторы, языковые эквилибристы, всевозможными способами нарушающие современные языковые нормы. Нарушение норм, конечно, лингвисту очень интересно (не случайно столько диссертаций защищалось, например, о языке Андрея Платонова), но нельзя заниматься только индивидуальными нарушениями нормы, оставив за пределами рассмотрения «настоящую» языковую среду. И опять же, стоит повториться, очень хороший баланс представляют в этом отношении мемуары: это тексты с элементами художественной организации, написанные, как правило, людьми незаурядными, хорошими рассказчиками, но при этом тексты, гораздо более укорененные в языковой стихии повседневности, чем современная художественная литература. Мой личный совет тем, кто хочет понять языковые вкусы и привычки какой-то определенной эпохи, — читайте мемуары, написанные свидетелями этой эпохи!

И вот, собранное вместе, всё это богатство, отражающее, как можно надеяться, современный русский язык после 1955 года во всех его аспектах, стилях, жанрах и формах существования (планируется разместить тексты общим объемом до 100 млн. словоупотреблений, сейчас — на момент подготовки сборника к печати — доступно около 70 млн<sup>3</sup>), позволяет любому человеку искать любые слова или сочетания слов в определенной грамматической форме или просто определенные грамматические формы. Говоря более техническим языком, возможен поиск по морфологическим параметрам. Кроме того, поиск возможен не только по всему корпусу, но и по определенному подмножеству текстов, выбранному пользователем: например, тексты определенного автора, определенного периода, определенного жанра и т. п. (естественно, в любых комбинациях).

Кому и зачем может быть нужен такой корпус?

---

<sup>3</sup> Отдельной частью проекта является включение в Корпус русских письменных текстов XIX и первой половины XX века. Эта работа также ведется; в настоящее время доступны тексты этого периода общим объемом около 15 млн словоупотреблений. Кроме того, на сайте Национального корпуса в настоящее время доступен экспериментальный корпус с синтаксической разметкой объемом около полумиллиона словоупотреблений и параллельный англо-русский и русско-английский корпус объемом около 2 млн словоупотреблений.

### Зачем нужен Национальный корпус русского языка?

Отчасти на этот вопрос ответ был уже дан. Прежде всего, корпус нужен — даже не просто нужен, а необходим — профессиональным лингвистам, тем, кто так или иначе имеет дело с фактами языка, а значит, должен эти факты собирать и систематизировать. Для лингвистов корпус — как минимум неопределимый инструмент, сокращающий затраты времени на техническую работу. На самом деле, конечно, корпус — нечто большее, чем просто техническое подспорье — это фактически справочно-информационная система по современному русскому языку, позволяющая получать ответы на самые неожиданные вопросы — более того, позволяющая ставить новые проблемы, которых лингвистика прошлого почти не касалась.

Несколько частных примеров, взятых из недавней исследовательской практики — моей или моих коллег. Вот на какие вопросы лингвист может ответить с помощью Корпуса буквально за считанные минуты:

1) В русском языке есть глагол несовершенного вида *реагировать*. Его коррелятами совершенного вида могут быть несколько разных приставочных глаголов: *прореагировать*, *отреагировать*, *среагировать* (явление нередкое, особенно среди заимствований). Какой из этих приставочных коррелятов употребляется чаще? К каким контекстам тяготеет каждый из этих приставочных коррелятов (например, какой из них охотнее сочетается с наречием *быстро*)? Наконец, в какой последовательности они появляются в современном языке — одновременно или по очереди? Различается ли частота их употребления в разные периоды?

2) В русском языке у части существительных мужского рода в парадигме единственного числа имеется особая дополнительная падежная форма — так называемый «второй родительный» падеж. Так, у слова *сахар* обычная форма родительного падежа выглядит как *сахара*, а форма второго родительного — как *сахару* (например, в контекстах типа *положите себе еще сахару*). Эта форма проникла в русское склонение в XVI-XVII вв., пережила «пик» употребительности в XVII-XVIII вв., а к концу XIX в. стала медленно угасать. В современном языке второй родительный встречается только примерно у сотни слов — правда, очень употребительных — и обычно может быть заменен на простой родительный (так, можно сказать и *положите себе еще сахара*, разница почти неощутима говорящими). Новые заимствования уже много десятилетий не при-



нимают формы второго родительного падежа. Можно посмотреть, с какой частотой употребляются у разных слов — и у разных авторов — формы второго родительного. Например, как сейчас предпочитают писать — *поднять с пола* или *поднять с полу*? А *встать* — *с пола* или *с полу*?

Совершенно ясно, что число таких примеров можно многократно увеличить, а задачи — усложнить. Но я нарочно привел примеры самых «обыденных» задач, встающих перед исследователями русского языка, чтобы показать, насколько Корпус технически сокращает и революционизирует работу с материалом даже и в этих случаях, позволяя простым нажатием кнопки получить данные, на сбор которых в иной ситуации потребовались бы месяцы (если не годы).

Сейчас использование корпуса в повседневной деятельности лингвистов постепенно становится нормой. Уже привычно, что доклады на научных конференциях сопровождаются словами «На основе примеров, полученных из Национального корпуса русского языка...» или «В Национальном корпусе русского языка это явление отражено следующим образом...». Студенты выполняют практические задания, пишут курсовые и дипломные работы, используя Корпус (и одновременно обучаясь правильно с ним обращаться). На материале Корпуса уже защищаются диссертации. Всё это совершенно естественно и могло бы произойти значительно раньше: в нашей стране корпусная лингвистика развивалась, к сожалению, с большим отставанием от США, Европы и Японии. Причин для этого, как всегда, было много, но будем надеяться, что сейчас этот разрыв начнет преодолеваться. Вполне возможно, что в недалеком будущем без корпуса изучаемого языка лингвисту будет так же невозможно обходиться, как, например, без словаря этого языка. Более того, корпус, словарь и грамматика, скорее всего, соединятся в один электронный ресурс — или базу данных, на основании которой и можно будет изучать язык.

Но, может быть, корпус — это изобретение ценное, однако интересное только узким профессионалам? Может ли корпус быть интересен кому-то еще, кроме лингвистов?

Безусловно, может — так же, как далеко не только лингвисты пользуются словарями и грамматиками.

Но сначала скажу о тех, кому давно уже не приходится специально рассказывать о преимуществах корпуса. Это — программисты

### **Зачем нужен Национальный корпус русского языка?**

сты, работающие в области автоматической обработки текстов (в том числе и различного рода поисковых систем). Поскольку программы такого рода имеют дело с естественным языком, они должны в той или иной степени «понимать» структуру текстов, написанных на этом языке. Причем — внимание! — понимать структуру именно тех текстов, которые встречаются в реальной жизни, а не идеализированную структуру языка в том виде, как она описана в академических и тем более школьных грамматиках. Из сказанного ясно, что программисты, может быть, как никакие другие профессионалы (не считая лингвистов, конечно) заинтересованы в том, чтобы корпуса, отражающие подлинное и максимально разнообразное языковое употребление, так сказать, неприглаженную языковую стихию, создавались и развивались. Не случайно проект создания Национального корпуса русского языка уже на самых ранних этапах был поддержан компанией «Яндекс» (крупнейший российский портал, разработавший одну из самых быстрых и эффективных поисковых программ по русскому сегменту Интернета), и в целом вне лингвистики наибольшее внимание к Корпусу проявляли и проявляют специалисты по информатике и программированию. Сотрудничество с программистами компании «Яндекс» оказалось очень плодотворным: при их непосредственной технической поддержке и был создан сайт, на котором размещен Национальный корпус русского языка в его нынешнем виде.

Разумеется, и программисты далеко не исчерпывают «целевую аудиторию» корпуса. Следующая по важности группа пользователей — это, наверное, те, кто так или иначе имеет дело с преподаванием современного русского языка, причем не только — и, может быть, даже не столько — в качестве родного, сколько в качестве иностранного.

Конечно, и в российской школе — в том числе и в высшей — Корпус может быть полезен для преподавателей и учащихся (об этом подробнее написано в статье Н. Р. Добрушиной в настоящем сборнике). Например, именно в Корпусе можно быстро и легко найти пример на редкое слово или грамматическую конструкцию, проверить написание или употребление интересующих единиц, и т. п. Следует, правда, при этом помнить, что Корпус — не нормативное пособие, язык там отражается не таким, каким он *должен быть* с точки зрения авторов описаний, а таким, каким он на самом деле *является*. С точки зрения школьной грамматики кое-что

из того, что встречается в Корпусе, будет квалифицировано как «отступление» от нормы. Впрочем, Корпус организован так, что учитель и его ученики имеют все возможности составить собственный подкорпус из образцовых с точки зрения языка произведений и пользоваться именно им.

Но многократно возрастает значимость Корпуса при обращении к иностранной аудитории. Для людей, не владеющих русским языком в качестве родного (как преподавателей, так и обучаемых), Корпус оказывается поистине незаменимым инструментом. Действительно, мы, носители русского языка, ежесекундно погружены в его стихию. «Примеры употребления» русского языка не просто существуют у нас в сознании — они окружают нас буквально повсюду. Конечно, специально отобранные и обработанные в составе Корпуса, эти примеры могут принести гораздо большую пользу, но если носителя русского языка лишить Корпуса, это не скажется фатальным образом на возможности его контактов со стихией русского языка — просто потому, что эта стихия существует по крайней мере и в его сознании тоже.

Иное дело иностранец. Его языковое сознание — не русское. И он в высшей степени нуждается в инструменте, открывающем ему максимально широкий (и максимально комфортный) доступ в мир русского языка. Ничего лучше Корпуса современная наука в этом случае предложить не может. Именно в корпусе преподаватель и студент могут найти ответы на многие интересующие их вопросы — причем такие ответы, которые и носитель не сразу догадается предложить. Поэтому не случайна высокая популярность корпусов в иноязычной среде. И именно от зарубежных русистов (в особенности преподавателей русского языка) мы получали самые заинтересованные и самые эмоциональные отклики.

В качестве небольшого отступления следует отметить, что первые электронные корпуса русского языка, доступные для всеобщего пользования, начали появляться не в России, а в Европе, и создателями их были европейские русисты. В свете сказанного это совершенно понятно. Специалисты хорошо знают так называемый Уппсальский корпус русского языка (включающий небольшой по нынешним меркам и практически не обработанный корпус русских художественных и публицистических письменных текстов 1960-х — 1970-х гг., с отдельными добавлениями из других периодов). Этот корпус был создан в Швеции, а в настоящее время, су-

### ***Зачем нужен Национальный корпус русского языка?***

щественно улучшенный и дополненный, хранится на сервере Тюбингенского университета в Германии, и в свое время немало пригодился специалистам. Интересные разработки по русской корпусной лингвистике ведутся в Финляндии — вообще, эта маленькая страна является одним из лидеров корпусной лингвистики в мире, в том числе и в области русской корпусной лингвистики. Весь этот опыт мы, конечно, по мере возможности старались учесть — наряду с нашим отечественным опытом последних десятилетий. (Подробнее о других электронных корпусах русского языка можно прочесть в статье Т. И. Резниковой и нашего финского коллеги М. В. Копотева в настоящем сборнике.)

Еще одна группа людей, для которых Корпус может представлять бесспорный интерес, — люди пишущие и, шире, вообще как-то в своей повседневной деятельности связанные со словом. Например, редакторы газет и журналов. Редакторам в своей практической деятельности гораздо чаще, чем академическим исследователям, приходится решать вопросы узуса: допустимо ли такое слово или конструкция? Кто, где, когда употребил впервые такой оборот? Для каких типов текста он наиболее характерен? Трудно представить себе, где можно быстро найти ответы на такие вопросы, если не в Корпусе. (Вообще, иногда кажется, что корпуса были придуманы специально для редакторов — впрочем, некоторые утверждают, что специально для программистов — или специально для преподавателей русского языка как иностранного.) Во всяком случае, нам известно, что многие, как теперь принято говорить, «продвинутые» редакторы активно пользуются Корпусом для решения своих повседневных проблем — ну и, конечно, для более эффективного устрашения авторов, как же без этого. «Вот Вы тут пишете... а между прочим, в Национальном корпусе русского языка...». Говорят, что такая синтаксическая конструкция действует безотказно. Опять-таки, Корпусом можно пользоваться и для проверки каких-то сведений, т. е. как справочником — но, еще раз повторю, в этом отношении необходимо соблюдать известную осторожность, помня, что не всё действительное может быть разумно (по крайней мере, если речь идет о тексте, предназначенном для публикации в газете или журнале).

Собственно, на этом можно было бы поставить точку, посчитав, что автор убедил читателей в необходимости пользоваться Национальным корпусом русского языка и обращаться к нему так же

регулярно, как и к словарю (если не сейчас, то в близком будущем). Однако в заключение, может быть, будет полезно поделиться некоторыми разрозненными наблюдениями, некоторыми оценками тех результатов, которые Корпус дает для изучения русского языка и понимания закономерностей его развития.

Выше мне уже приходилось говорить, что одна из наиболее интересных задач, которую можно решать именно с помощью Корпуса — это наблюдения за динамикой развития языка. Поскольку тексты, входящие в корпус, датированы, то нетрудно проследить за хронологией постепенных языковых изменений — за появлением или постепенным угасанием определенных слов, конструкций или грамматических форм (типа «второго родительного» падежа).

Это вызывает к жизни фактически новое направление — своего рода «микроисторическую» лингвистику, в центре внимания которой находятся не глобальные изменения в истории языка, а изменения менее масштабные, занимающие десятилетия (для истории языка это чрезвычайно маленький срок). Нельзя сказать, что такие исследования раньше не проводились вовсе (напротив, их очень любил, например, один из классиков отечественной филологии академик В. В. Виноградов), но с появлением корпусов их объем и эффективность, по моему мнению, должны стремительно возрасти.

Что же говорит нам Корпус о современном состоянии русского языка? Опуская специальные подробности, можно заметить следующее.

С точки зрения истории русского языка, нынешний период является чрезвычайно важным и интересным. Это — период языкового сдвига, может быть — языкового слома. Вообще в истории языка периоды относительной стабильности всегда чередуются с периодами относительно быстрых и глубоких изменений. Можно полагать, что русский язык вплотную приблизился к такому состоянию, и вероятность больших изменений в его структуре в ближайшем будущем весьма велика.

Об этом косвенно свидетельствует то, что предыдущие столетия в целом можно охарактеризовать как относительно стабильные. Последний «неспокойный» период для русского языка — это XVII-XVIII в., когда после петровских реформ в язык хлынул поток европейских заимствований, старославянский язык был значительно потеснен в письменном узусе и завершился переход от среднерусской системы к ранней современной русской.

### Зачем нужен Национальный корпус русского языка?

Нынешнюю историческую эпоху часто сравнивают именно с этим временем. Резкий слом привычных социальных условий. Изменение самого статуса литературного языка. Стилистический, жанровый (и даже орфографический, совсем как в те времена!) разнотой. Поток заимствований (в нынешней ситуации — в основном английских). Какими же могут быть результаты? Не следует ли говорить, что в начале XXI века складывается новая языковая система, по отношению к которой нынешняя будет называться как-нибудь вроде «поздний среднерусский» или «поздний классический период»?

Конечно, сейчас такие прогнозы делать рано, но изменения в самом строе русского языка весьма вероятны. Если более 200 лет с языком «почти ничего» не происходило, то критический потенциал изменений возрастает. Нужен лишь внешний толчок, катализатор.

Интересно, что резких изменений в русском языке многие ожидали в начале XX века, после революции. И, действительно, ситуация начала 1920-х тоже во многом напоминала нынешнюю. Но тогда эти ожидания не оправдались: русский язык пережил социальные потрясения сравнительно безболезненно. Глубинные пласты языковой системы почти не были затронуты изменениями. Если вдуматься, это факт и удивительный, и в то же время закономерный: советское общество очень быстро — через 10-15 лет после революции — позиционировало себя как общество закрытое, жестко регламентирующее все сферы жизни (в том числе и письменный язык!) и при этом консервативное. Такая политика очень эффективно «заморозила» ростки языковых изменений, которые могли бы уже в то время дать весьма непредсказуемые побегі. В результате современный русский язык гораздо ближе стоит к языку начала XX века, чем, например, современный французский, стремительные изменения которого ничем (кроме обычных общекультурных механизмов современного государства) не сдерживались. Но вторая попытка — после новой революции — может оказаться более действенной.

Из конкретных вещей нужно отметить, что, как всегда, лексика и фонетика идут впереди. Лексический состав русского языка сейчас меняется стремительно как никогда: иноязычные заимствования, сленг, внутренние изменения значений слов (вспомним, что теперь на самом деле значат такие слова, как *напрягать* или *грузить*) — всё это делает свою ежедневную муравьиную работу. Ме-

**В. А. Плунгян**

---

няется и произношение: значительно ускорился темп речи, проглатываются интервокальные согласные («хоит» вместо *ходит*, «буиш» вместо *будешь* — и даже «поинл» с особым «гнусавым» призывком вместо *понял*). Почти не слышны безударные гласные на конце слова.

Рано или поздно это должно сказаться на слабых зонах русской грамматической системы. Грамматика пока держится. Я думаю, наиболее уязвимо в русском языке склонение. Тексты Корпуса показывают заметный рост — особенно в устной речи — именительного падежа во всех позициях. Так, числительные уже практически перестали склоняться (совсем не трудно услышать что-нибудь вроде *остался без триста двадцать пять тысяч рублей*). На наших глазах лишаются форм склонения и географические названия типа *Бородино*: в повседневной речи фактически единственным вариантом становится *рядом с Бородино, живу в Бородино* и т. п. Не желают склоняться, что характерно, и новые заимствования: а заимствования никогда не подчиняются тем законам, которые в самом принимающем языке действуют слабо. Ждет ли русский язык судьба болгарского — потерявшего все падежи, функции которых взяли на себя предлоги? Скорее всего, всё-таки нет. Но «что-то может произойти».

Теперь у нас есть Корпус. Так что получить ответ, может быть, будет проще.