

ИССЛЕДОВАНИЯ

В. А. ПЛУНГЯН

КОРПУС КАК ИНСТРУМЕНТ И КАК ИДЕОЛОГИЯ: О НЕКОТОРЫХ УРОКАХ СОВРЕМЕННОЙ КОРПУСНОЙ ЛИНГВИСТИКИ

*У нас дома говорили по-русски чисто
и правильно, и корпусные выражения мне
резали слух.*

Гайто Газданов. Вечер у Клэр (1930)¹

И. В настоящих заметках речь пойдет в основном о вещах достаточно абстрактных и общих, касающихся тенденций развития современной лингвистики. Однако все наши рассуждения будут иметь совершенно конкретный источник — это недавно созданный Национальный корпус русского языка (с апреля 2004 г. доступный в интернете по адресу www.ruscorgo.ru); в настоящее время он насчитывает более 150 млн. словоупотреблений и включает (в различной пропорции²) тексты с начала XVIII по начало XXI в. Собственно описание корпуса, а также возможностей его непосредственного практического применения не является нашей задачей — это уже было сделано в ряде публикаций последнего времени (см. прежде всего сборник [НКРЯ 2005], а также статьи [Резникова, Копотев 2005] и [Резникова 2008], где проводится сравнение Национального корпуса русского языка с существующими корпусами русского и других славянских языков). Мы хотели бы коснуться несколько иного аспекта использования корпуса — так сказать, идеологического. Дело в том, что появление корпуса не просто дало в распоряжение лингвистов новый мощный инструмент анализа фактов языка — оно в определенной степени изменило теоретические приоритеты и отчасти даже взгляды на

¹ Пример получен, разумеется, из Национального корпуса русского языка.

² Наиболее подробно и разнообразно представлен современный период, т. е. тексты второй половины XX — начала XXI в.; этих текстов не только больше количественно, но они включают и такие жанры и типы, которые в других временных периодах представлены ограниченно или вовсе не представлены (например, записи устной речи, сценарии кинофильмов, личные письма и дневники, газетные объявления).

то, чем является язык и какие задачи изучения языка являются наиболее важными.

Поскольку частично эти изменения приоритетов оказались созвучны тем тенденциям, которые и до этого стали обозначаться в развитии теоретической лингвистики достаточно отчетливо, целесообразно напомнить, каковы эти тенденции.

Как кажется, сейчас в мировой лингвистике намечаются контуры новой модели языка, которая в ряде существенных отношений отличается от привычных моделей, сложившихся к последней четверти XX века. Говоря об этих моделях предшествующего этапа, часто можно, как ни странно, отвлечься от различий между разными научными школами — эффект «исторического расстояния» приводит к тому, что те направления, которые казались (и считали себя) непримиримыми антагонистами, на самом деле обнаруживают значительно больше общих черт, чем различий (таково, например, соотношение структурных и ранних генеративных моделей, и т. п.). Тем самым можно, по-видимому, в некоторых случаях апеллировать к некой обобщенной лингвистической идеологии «середины XX века», противопоставленной идеологии нынешнего рубежа столетий (следует подчеркнуть, однако, что в таких случаях всегда имеется в виду именно господствующая или преобладающая идеология, поскольку «теоретические диссиденты» встречались на каждом этапе существования лингвистики — и иногда их взгляды становились преобладающими позднее). С другой стороны, те тенденции, о которых пойдет речь ниже, сейчас отчетливее проявляются в тех современных теориях языка, которые относятся к так называемому функциональному (или функционально-когнитивному) направлению; может быть, просто потому, что они менее консервативны, возникли позже, чем «формальные» теории, и во многом как полемика реакция на них. Но при этом и функционализм «докорпусной» эпохи тоже в незначительной степени отличается от современного функционализма.

Если попытаться обозначить общее отличие современных моделей языка от моделей предшествующего периода, то одной из наиболее существенных черт будет, по-видимому, отчетливая смена теоретических приоритетов — с переходом от «системы» к «узусу» и от «языка» к «речи». Этот переход посягает на, казалось бы, незыблемый постулат лингвистических теорий, вошедший во все учебники языкознания: на фундаментальное различие между языком и речью и на обязанность лингвистики быть в первую очередь наукой о языке (или «языковой компетенции»), а не наукой о речи (или «речевом узусе»).

Конечно, здесь проявляется и вполне понятное «маятникообразное» развитие науки, когда господствующая теория себя исчерпывает и ее недостатки начинают заслонять ее достижения. Современные исследователи стали всё больше призывать лингвистику изучать факты, а не конструкты; рассуждать о свойствах наблюдаемых явлений, а не о свойствах моделей. Проблема же лингвистики в том, что язык, как известно, не наблюдаем и

представляет из себя, в некотором смысле, теоретическую абстракцию. Наблюдаемыми являются речь, т. е. процесс построения текстов, и, разумеется, сами тексты. Лингвистам со студенческой скамьи объясняли, что, конечно, они видят перед собой тексты, но не тексты являются для них по-настоящему важными: минувя их, они должны перейти к более значимому объекту — к системе правил, по которым эти тексты строятся, — и рассуждать об этой системе, а не о текстах, «реконструируя» ее на основе текстов. Положение текстов в такой идеологии самое незавидное: они находятся в области эмпирического, случайного, внесистемного. Существовая на самом деле (в отличие от «системы языка»), они как раз и объявлялись как бы несуществующими для теории.

В таком отношении к текстам заключалась, конечно, большая несправедливость. И всё же следует отдать должное теоретикам языка. Пройдя очень далеко по пути отрицания самостоятельной значимости «речевого» или «текстового» материала, многие из них в конце концов остановились и задумались: а существует ли язык вообще? Не является ли эта абстракция слишком сильной? И не пора ли вернуться назад, к рассмотрению именно текстов, причем реальных текстов, которыми люди обмениваются в процессе коммуникации, а не искусственных предложений на страницах лингвистических статей?

Именно этот вектор развития можно считать ответственным за те подходы (и те сдвиги в выборе приоритетов), которые очень отчетливо проявили себя в последние десятилетия. Можно обозначить следующий список «идеологических предпочтений», характерных для этого нового взгляда на задачи теоретической лингвистики:

1) Внимание не к слову или к предложению, а к тексту, или, как теперь чаще говорят, к **дискурсу** — то есть к реальному инструменту коммуникации в целом, а не к его отдельным фрагментам. Эта тенденция обозначилась еще в 1970-е гг., но в последнее время проявляется особенно интенсивно.

2) Внимание к **квантитативному** компоненту языка, т. е. учет в первую очередь более частотных (в дискурсе!) элементов по сравнению с менее частотными, и признание квантитативных отношений существенным фактором в языковой эволюции и структуре языковых правил. В западной лингвистике это, пожалуй, особенно характерно для школы Дж. Байби (ср., например, [Вубее 2001; 2006]), хотя проявляется и у многих других исследователей; элементы этого подхода мы находим ещё у Гринберга в 1960-е гг. (с его теорией маркированности) и др., но тогда они были периферийны³.

³ Можно вспомнить и следующую цитату из Н. С. Трубецкого, звучащую удивительно современно: «В связи с работой над морфологией и морфонологией прихожу к убеждению, что необходима и морфологическая статистика. Очень важно знать, какие категории употребляются чаще и какие реже» (из письма к Р. О. Якобсону от 29 окт. 1932; см. [Трубецкой 2004: 260]).

3) Внимание к **синхронной вариативности** языка, т. е. признание того факта, что в рамках данной языковой общности не существует единой жесткой системы средств выражения смысла, а существуют различные стратегии реализации этой задачи, в том числе зависящие от психологических, биологических и социальных факторов (но не только от них). Эта тенденция также обозначилась еще в середине XX в. с возникновением социолингвистики и других дисциплин этого круга.

4) Внимание к **диахронической вариативности** языка, т. е. признание того факта, что язык постоянно изменяется во времени и полностью отвлечься от этой нестабильности невозможно, что в каждый момент времени в языке сосуществуют «прогрессивные» и «консервативные» участки. Тем самым признается, что «строго синхронное» описание языка является иллюзией; более того, полное грамматическое описание языка должно иметь своего рода «размытый» характер, включая сведения не только о способах выражения грамматических значений, но и о динамике изменений этих способов на протяжении наблюдаемых исторических отрезков (одно-два столетия). Эти идеи также не являются новыми — они высказывались многими лингвистами середины XX в. (в том числе уже упоминавшимся выше Дж. Гринбергом, ср., например, его известную статью [Greenberg 1979])⁴.

5) Изменение отношения к понятию языковой нормы и языковой правильности: возобладало, так сказать, более толерантное отношение к этим понятиям. Граница между «ошибкой» и «маргинальным вариантом», а также между маргинальным вариантом и полноценным (а впоследствии, может быть, и единственным) признается гораздо более подвижной и зыбкой. Исследователь предпочитает помнить, что сегодняшняя ошибка вполне может оказаться завтрашней нормой, и воздерживается от поспешных и тем более оценочных суждений.

Эти (как и ряд других) установок в современных работах чаще всего объединяются под названием «подход, ориентированный на узус» или «теория узуса» (англ. *usage-based approach* — термин, предложенный Р. Лэнгакером в начале 1990-х гг.). Эта модель представлена в работах таких (в остальном) разных американских и европейских лингвистов, как типологи и теоретики функционализма Дж. Байби, Т. Гивон и У. Крофт, пси-

⁴ В этом контексте заслуживает упоминания относительно малоизвестная у нас статья французского слависта Поля Гарда [Garde 1988], содержащая обоснование так называемого «бисинхронного» подхода к анализу языковых явлений. Этот подход предусматривает описание фактов языка не в рамках единой системы «правил» и «исключений», а в рамках двух сосуществующих систем — более старой и более новой. В статье Гарда в качестве примеров рассматриваются, в частности, правила акцентуации презенса русских глаголов и синтаксис русских числительных. Удачное применение этой же модели к анализу предикативных употреблений русских прилагательных предложено в работе [Гиро-Вебер 1996].

холингвист М. Томаселло, славист Л. Янда и мн. др. (у нас во многом близкий подход представлен в ряде последних работ А. Е. Кибрика). Теория узуса стремится поставить в центр изучения и теоретического анализа непосредственно дискурсивную практику; несколько упрощая, ее положение можно сформулировать так: не существует «языка вообще», а существуют структуры, которые преобладают в определенных типах дискурса (у разных говорящих, в разные моменты времени и т. п.); их и надо изучать.

Подход, ориентированный на узус, противостоит «системному» подходу, ориентированному на изучение некоторой идеальной структуры языка, по отношению к которой наблюдаемые факты являются лишь более или менее адекватными реализациями (и если какие-то реализации исследователь сочтет неадекватными, он вполне может их проигнорировать). Системный подход, возникший в рамках структуралистского этапа развития лингвистики, был полностью унаследован современными «формальными» синтаксическими теориями. Идеологические различия двух подходов определяют и различное отношение их представителей к языковым фактам. Для сторонников системного подхода характерно недоверие к «экспериментальным» и «объективным» методам исследования материала, которые, с их точки зрения, мешают увидеть столь любимые ими «обобщения» (как правило, сформулированные уже до начала всякого исследования), затемняя их ненужными эмпирическими случайностями. Таким исследователям, как правило, вполне достаточно собственной интуиции: и действительно, если целью исследования является не учет вариативности, а ее устранение в пользу некоторой идеальной картины стройной и статичной системы, то автор-лингвист не только не хуже, но даже предпочтительней других носителей языка — он гораздо лучше знает, каким язык должен быть. Напротив, сторонникам «теории узуса» крайне важна эмпирическая база, т. е. представительная совокупность текстов на данном языке: тексты являются их реальным объектом, свойства которого не известны до начала исследования и именно в ходе исследования и могут быть обнаружены. Исследователи этого типа хорошо понимают, насколько обманчивой может быть интуиция любого отдельного носителя (в том числе и лингвиста — а может быть, вернее сказать, и в особенности лингвиста). Точно с тем же недоверием, с каким сторонники системного подхода относятся ко всякого рода «экспериментальным проверкам», сторонники теории узуса относятся к суждениям, основанным на не подтвержденной фактами интуиции: язык шире любого отдельного носителя, тем более носителя пристрастного, т. е. озабоченного поиском некоего абсолюта — будь то подтверждение принципов «универсальной грамматики» или, например, свидетельств преобладания «настоящей московской нормы».

Итак, исследователи, опирающиеся на «узус», склонны для подтверждения своих гипотез обращаться к представительной совокупности текстов на данном языке. А представительная совокупность текстов и является де-факто тем, что в современной лингвистике принято называть словом

«корпус». Тем самым, ключевое слово настоящих заметок возникает естественным образом как следствие изложенных выше «новых» теоретических предпосылок — дискурсивных, квантитативно-динамических, эмпирических и, если угодно, антисистемных. Не удивительно, что эта новая лингвистика очень быстро оказалась лингвистикой корпусов. В каком-то смысле трудно сказать, что здесь причина, а что следствие: то ли идеологические установки «теории узуса» (начавшей формироваться, как мы уже отмечали, никак не позднее последней трети XX в.) привели к бурному расцвету корпусных исследований, то ли прогресс в такой первоначально сугубо прикладной области, как составление электронных корпусов языков, привел к резкой смене идеологических установок лингвистов (или ускорию эту смену). Скорее всего, имели место оба процесса, и правильнее говорить о встречном движении: в ходе теоретических поисков был обнаружен наиболее подходящий для этих поисков инструмент.

Подчеркнем еще раз, что в настоящее время корпус — это не просто дань техническому прогрессу или более удобный инструмент для поиска примеров; это именно примета новой идеологии изучения языка, для которой язык — вообще говоря, *и есть корпус*, как бы вызывающе это ни звучало по отношению к предшествующей теоретической традиции. Таким образом снимается одно из главных возражений сторонников системного подхода, которые всегда любили настаивать на «бесконечном», «открытом» характере языка по сравнению с «конечной» совокупностью любых текстов на данном языке. Однако, с одной стороны, нельзя не видеть, что современные корпуса предоставляют исследователю настолько большие массивы текстов, что для отдельного человека они практически оказываются бесконечными, а с другой стороны, нельзя забывать и об очень плодотворном опыте изучения мертвых языков, «конечность» которых никак не сказывается на качестве их описаний (а если сказывается, то часто едва ли не в лучшую сторону). Дело скорее в том, что раньше в нашем распоряжении просто не было технических средств, позволяющих получить быстрый и эффективный доступ к по-настоящему значительной совокупности текстов, а сейчас такие средства есть. Тексты поневоле казались вторичным и конечным объектом, так как оценивалась только та их совокупность, которая была соизмерима возможностям отдельного, медленно пишущего и медленно читающего человека. Реальный объем текстов на любом живом языке на много порядков превосходит эту совокупность. И только сейчас стало по-настоящему понятно, насколько за предыдущие десятилетия лингвистика успела оторваться от своего главного объекта — текстов, и насколько многие утверждения, сделанные в рамках разных теорий, не подтверждаются фактами корпуса (в отношении русского языка красноречивые примеры такого рода приведены в недавних публикациях [Перцов 2006а; 2006б]; некоторые другие примеры мы обсудим ниже). Даже для тех исследователей, которые изначально были ориентированы не на интроспекцию, а на обработку больших массивов данных, корпус оказыва-

ется неоценимым инструментом, во много раз повышающим скорость и надежность работы. Здесь наблюдается своего рода двойной эффект. До появления электронных корпусов текстов возможности ручной обработки материала были относительно скромные: на поиск нужных примеров часто уходили месяцы, при этом трудно было дать гарантию того, что самые важные контексты действительно оказались учтены. С появлением корпуса даже те лингвисты, которые понимали важность обращения к такого рода материалу, получили результат, во много раз превзошедший их ожидания.

Интересным образом, описываемый подход в некоторых чертах сближается с традиционной «деструктуралистской» филологией, хотя, конечно, не повторяет ее установки полностью. Это достаточно характерная тенденция — современная функциональная лингвистика вообще, как известно, во многом вернулась к тому, что декларировалось в XIX веке, а потом было забыто или отброшено. В интересующем нас аспекте можно вспомнить классическую филологию, всегда имплицитно исходившую из того, что, например, классический латинский язык — это не больше (но и не меньше) чем корпус текстов определенного периода. Вообще, как уже было сказано выше, хорошее описание мертвого языка — это максимально тщательное описание всей совокупности текстов, входящих в корпус этого языка (именно так устроено, например, предложенное А. А. Зализняком [2004] описание языка древненовгородских берестяных грамот). Но к корпусным по сути методам тяготеют не только исследования мертвых языков — сходным образом строятся, например, работы такого известного в традиционной филологии жанра, как исследования «языка писателя», когда материалом служит корпус произведений одного автора и всё, что представлено в таком корпусе, равно заслуживает внимания. Традиции «виноградовской школы», столь популярные в отечественной русистике, — это во многом традиции «стихийной» корпусной лингвистики, ориентированной на представительные (и, как правило, закрытые) совокупности литературных текстов писателей, входивших в литературоведческий канон.

По последнему поводу, впрочем, уместно заметить, что у современной корпусной идеологии (в том числе и реализованной в рамках Национального корпуса русского языка) есть и черта, отличающая ее от подходов, принятых в традиционной филологии. Это — принципиальная «нелитературоцентричность», характеризующая отбор текстов для корпуса. Конечно, роль текстов, представляющих классическую и современную художественную литературу, в корпусе достаточно велика, но существенно, что, в отличие от традиционных описаний, учет именно этих текстов не является для многих задач приоритетным. В особенности это относится к современной постмодернистской художественной литературе, отказавшейся как от претензий на общественную значимость (по крайней мере, если понимать последнюю как публицистичность «прямого действия»), так и на языковую нормативность, т. е. апелляцию к повседневной языковой практике «среднего образованного человека». Взамен же эта литература приобрела уста-

новку на языковую игру, на извлечение художественного эффекта из многообразных нарушений нормы; тем самым порождаются экспериментальные тексты, сами по себе не лишенные интереса, но никак не способные служить образцом доминирующего в данном языковом коллективе дискурса. На роль последнего может в современной ситуации претендовать скорее литература, относимая к жанру 'non-fiction', то есть литература с минимально декларируемой «художественностью», а также образцы устного городского фольклора: анекдоты, анонимные «истории из жизни», вербализующие стереотипы и мифы современного массового сознания, и т. п.

II. Подводя итоги сказанного, полезно еще раз подчеркнуть, что в современной теоретической лингвистике корпус — это не только мощный инструмент исследования языка, но и новая идеология, ориентирующая исследователя на *текст* как главный объект теоретической рефлексии. Можно сказать, что корпус в каком-то смысле вернул лингвистам их подлинный объект — тексты на естественном языке в максимально полном объеме, и последствия этого неожиданного приобретения, на наш взгляд, еще скажутся в ближайшем будущем не только на лингвистической практике, но и на лингвистической теории. Таким образом, с помощью корпуса стало возможно не только быстрее и эффективнее решать известные науке задачи, но и ставить принципиально новые задачи, ранее практически невыполнимые из-за их трудоемкости. К последнему типу задач относятся прежде всего всевозможные обобщения по поводу «микроэволюции» языка на протяжении одного-двух столетий: малозаметные изменения сочетаемости и значений слов, изменения частотности различных конструкций или частотности употребления лексических и грамматических вариантов, регистрация появления или угасания отдельных явлений языка, и т. п.

Собственно, можно надеяться на то, что будущие грамматические описания языков станут «корпусно-ориентированными» — в том смысле, что всякое утверждение про данный язык, которое будет в них делаться, можно будет проверить относительно некоторого корпуса этого языка. Это, безусловно, повысит научный статус таких грамматик, поскольку количество эмпирически проверяемых утверждений в них резко возрастет по сравнению с текущим состоянием, когда грамматики в основном имеют дело либо с отдельными текстовыми примерами (которые могут иллюстрировать описываемое явление, т. е. являться доказательством его существования в языке, но не могут использоваться для исследований его частотности), либо с искусственно сконструированными исследователем языковыми выражениями.

Вообще, использование корпуса особенно тесно связано именно с проблемой доказательства существования в языке того или иного явления. Конечно, нельзя полностью отождествлять понятия «иметься в корпусе» и «иметься в данном языке». В языке могут существовать потенциально возможные явления, не отраженные в конкретном корпусе текстов, хотя в

случае большого представительного корпуса сам факт отсутствия некоторого явления (пусть и потенциально возможного) всё равно значим. С другой стороны, в корпусе могут встретиться окказиональные явления, да и просто ошибки, которые будут отвергнуты говорящими на данном языке при предъявлении им соответствующих образцов для оценки правильности / коммуникативной уместности. Однако частотность таких явлений — если они действительно окказиональны — будет крайне мала. Вместе с тем «ошибка» (и, шире, любое отступление от того, что принято считать «нормой»), систематически фиксируемая в корпусе, — возможно, уже не ошибка и требует более внимательного к себе отношения. Известно, что авторы нормативных пособий (как и лингвисты, более всего опирающиеся на интроспекцию) нередко склонны выдавать желаемое за действительное и описывать не столько язык, сколько свои представления о том, каким он должен быть (подробнее на эту тему см. также упомянутую выше работу [Перцов 2006б]).

В статьях Н. В. Перцова критике подвергаются, в основном, суждения о приемлемости или неприемлемости тех или иных русских конструкций, делавшиеся специалистами по лексической семантике. Однако сказанное выше о том, что «язык шире любого отдельного носителя», даже в большей степени применимо и к работам современных синтаксистов, постулирующих те или иные синтаксические ограничения. Так, в теории синтаксиса одним из критериев выделения составляющей считается ее способность замещаться анафорическими единицами — так называемыми «проформами»; эта способность имеется у составляющей в целом, но не имеется у отдельных ее элементов: «⟨...⟩ терминальные составляющие ⟨...⟩ не заменяются на проформы: это возможно только для фразовых категорий» [Тестелец 2001: 142]. В учебнике Я. Г. Тестельца именно этот тест приводится в качестве одного из доказательств существования именных групп: и сочетания типа *этот человек*, и сочетания типа *этот человек, который...* допускают замену на *он* только целиком, так как ни конструкции типа *этот он*, ни конструкции типа *этот он, который...* невозможны.

На первый взгляд это утверждение кажется абсолютно очевидным. Тем не менее, в корпусе русского языка обнаруживается немало примеров, не согласующихся с ним. Не говоря уже о конструкциях типа *он, который сделал для нас так много...*, представленных многими десятками употреблений (ср. 1–2), обнаруживаются, что самое удивительное, и разнообразные примеры конструкций вида *этот он, который...* (ср. 3–5):

- (1) *Шепотом вскрикивал, что он **ее, которая** толкала его на борьбу, ничуть не винит, о нет, не винит!* (М. Булгаков. Мастер и Маргарита, часть 1 (1929–1940))
- (2) *Мне не жаль было ее умельца-мужа, для которого тысяча «горбачих» — как раз столько, сколько мы с женой тратили на весь не-*

дельный хлебный паек, — были не деньги. Не жаль было и **ее, которая** сейчас с сотней, а то и двумя этих тысяч приехала «по обуви» и, вспоминая мужа, тащится со мною ночью на площадь. (Ал. Кабаков. Путешествие экстраполятора (1988–1999))

- (3) Он сел, выпрямил ноги. Большая красная волна подняла его и опять опустила. Сидя, он видел себя, лежащего. У **того его, который** сидел, ноги были выпрямлены, а у того, который лежал, — поджаты. (И. Грекова. Фазан (1984))
- (4) Черный ангел прилетел. Пока тот не исчез, Мокрухин протянул руку и втянул ангела в квартиру, и Евгения краем сознания почувствовала, что **та она, которая** стояла у лифта, медленно двинулась вслед. Дверь закрылась. (Елена и Валерий Гордеевы. Не все мы умрем (2002))
- (5) Данка чувствовала, что с каждым шагом приближающегося к ней Термосёсова покидают ее последние силы. Она не знала, что он скажет, что сделает, вообще с чего начнет и на чем станет? И, наконец, на чем может остановиться он, **этот он, который** от первой минуты своего появления до этого решительного заключения на замок судьбы, ни на минуту не перестает изумлять ее? (Н. С. Лесков. Божедомы (1868))

Конечно, такие конструкции встречаются редко; конечно, их существование не очевидно даже искусственному лингвисту. Но отрицать возможность таких конструкций никак нельзя — а следовательно, утверждения о синтаксических критериях для выделения составляющих всё-таки нуждаются в определенной коррекции. Убедиться в этом удастся лишь благодаря апелляции к корпусу.

С учетом сделанных оговорок, можно, как представляется, отождествить понятия «существующего в языке» и «надежно засвидетельствованного в корпусе». В известной дихотомии системы и узуса корпус ориентирует лингвистов на узус — хотя бы по той простой причине, что никогда раньше узус не был так хорошо доступен для исследований. Между тем, именно узус — т. е. тексты — и является единственной подлинной реальностью науки о языке, т. е. объектом, доступным непосредственному наблюдению.

Идеология корпусной лингвистики позволяет сделать и более сильное утверждение: при прочих равных условиях, должно быть предпочтительнее такое описание языка, которое будет ориентировано на явления, лучше представленные в корпусе. Иными словами, то, что является центральным в корпусе, должно быть центральным и в грамматике. Плодотворность для теории языка описания механизмов выражения таких смыслов, которые не являются востребованными реальными говорящими на данном языке, как минимум неочевидна. Как кажется, по крайней мере полезнее вначале опи-

сать способы выражения тех смыслов, которые для данного языкового коллектива являются приоритетными, — то есть те, которые легко найти в корпусе данного языка. Язык прежде всего дает нам возможность выразить и понять то, что уже было многократно сказано в данном национальном пространстве; может ли язык, без существенного насилия над его структурой и механизмами, быть успешно использован как-то еще — вопрос, вообще говоря, открытый. Может быть, и может — и если так, то такие смещенные способы использования языка тоже должны быть описаны. Но здесь существенны приоритеты, и использование корпуса, как кажется, позволяет расставить эти приоритеты достаточно надежно. Смысл корпусно-ориентированной лингвистики в том, что она позволяет изучать действительно существующие в языке, а не мнимые явления⁵.

III. В заключение остановимся кратко на том аспекте описания современного русского языка, который автору этих строк в силу обстоятельств наиболее близок, а именно, на некоторых следствиях для описания русской морфологии, которые возникают при попытке посмотреть на эту область в зеркале корпуса.

Заметим сразу, что даже такая сравнительно консервативная область языка, как морфология, демонстрирует достаточно много существенных отличий от той картины, которая представлена в нормативных описаниях.

⁵ Как известно, взгляды современной генеративной лингвистики на эти проблемы едва ли не полностью противоположны. Считая основным объектом описания «языковую компетенцию» некоторого идеального говорящего, теоретики этого направления по-прежнему полагают, что основным инструментом описания этой компетенции являются суждения относительно «грамматической правильности» сконструированных высказываний, причем чем дальше эти высказывания будут от возможных в естественных текстах, тем это лучше для теории: таким образом проверяется именно врожденная языковая компетенция, не замутненная знакомством с реальными образцами высказываний на данном языке. Ср. характерные утверждения П. Постала, сделанные им еще в известной монографии 1974 г., в защиту использованных там крайне громоздких искусственных английских примеров, и сочувственно воспроизводимые в одной из новейших работ: «⟨...⟩ the linguistic experience of speakers will not dependably provide them with the opportunity to come in contact with the relevant examples. ⟨...⟩ Consequently, the definite judgments available to English speakers about such cases must follow from general principles that are internalizable independently from such marginal sentences» [Postal 1974]; воспроизводится в [Davies & Dubinsky 2004: 46] (благодарю П. М. Аркадьева, любезно обратившего мое внимание на эту цитату). Подобная аргументация может звучать убедительно только в одном случае: если полностью забыть, что, за отсутствием мифического «идеального говорящего», грамматическую правильность подобных примеров приходится на свой страх и риск оценивать скромным авторам проверяемых теорий. О том, что происходит при предъявлении сконструированных генеративными синтаксистами примеров обычным носителям языка, подробно рассказывается, например, в остроумной статье Евы Домбровской [Dąbrowska 1997].

Ведь корпусное описание (в том числе и в силу необходимости обеспечить корректный автоматический анализ текста) не может уклониться от того, чтобы анализировать *все* встречающиеся в текстах формы. А таких форм заведомо больше, чем это предписывается нормативной грамматикой. Причем, что существенно, в число примеров русского «морфологического расширения» попадают далеко не только случайные ошибки или окказиональные употребления — это и частотные разговорные формы, пока (почти) не находящие отражения в грамматиках, это и частые случаи языковой игры, также имеющие свою «грамматику» и свои регулярные механизмы⁶, это и такие устаревшие, диалектные и просторечные формы, которые, несмотря на исключение их из нормативных грамматик, продолжают существовать в языке, известны говорящим и могут употребляться в соответствующих контекстах.

Но в то же время реально встречающихся в текстах форм в каком-то смысле и меньше, чем в нормативных источниках. Существующие системы автоматического морфологического анализа, как правило, не ориентированы на частотность той или иной формы: если форма в принципе возможна, она присутствует в автоматическом словаре. Следствием такого подхода оказывается то, что порождается немалое количество форм, которые, хотя и возможны по правилам русской морфологии, в реальных текстах никогда не встречаются. Это «гиперпорождение» не всегда безобидно: специалистам по автоматическому морфологическому анализу хорошо известен эффект паразитической омонимии, затрудняющий работу программ; этот эффект возникает в тех случаях, когда у распространенной словоформы появляется мифический двойник, формально законный, но реально в текстах отсутствующий. К таковым, например, относятся разбор предлога *для* как деепричастия от глагола *длить* или разбор формы презенса *сеем* как краткой формы страдательного причастия *сеемый* (даже полная форма этого причастия встречается в корпусе всего один раз); проблематично и порождение форм множественного числа у слов типа *современность*, и мн. др.

Таким образом, морфология, отражающая данные корпуса, одновременно и шире, и уже «стандартной» — она включает в себя в качестве легитимных многие не предусматриваемые стандартной грамматикой варианты, но зато исключает надежно прописанные в стандартной грамматике «потенциальные» (т. е. возможные лишь теоретически) формы.

⁶ А в современных текстах к этим случаям добавляются и стремительно конвенционализирующиеся приемы орфографической игры (с написаниями типа *нраввицо* или *цаз*), так что любому исследователю, имеющему дело с автоматической обработкой текстов современного русского сегмента интернета, нельзя не считаться с этой «новой» орфографической вариативностью — как, впрочем, и с «обычными» случаями неустойчивой орфографии в недавних заимствованиях (*плеер*, *плейер* и *плэйер*, *хэнд-аут* и *хендаут*) и слэнговых словах, редко попадающих в кодифицированные письменные тексты (*голимый* и *галимый*; *децл*, *дэцл*, *децил*, *дэцэл*) и т. п.

Из числа наиболее интересных морфологических явлений, бесспорно существующих в современных русских текстах, но недостаточно отражаемых нормативной грамматикой (или вовсе не отражаемых), можно упомянуть следующие:

1) Появление особых «стяженных форм» винительного и дательного падежей у личных местоимений *я* и *ты* (*мя, тя, те*), по своим морфосинтаксическим свойствам отчасти напоминающих древнерусские местоименные энклитики; такие формы (существовавшие в диалектах и в просторечии) не только надежно фиксируются в устной речи всех слоев носителей русского языка, но и начинают в заметном количестве проникать в некодифицированную письменную речь (особенно в тексты электронной коммуникации); к этому же классу примыкает и частотная форма именительного и винительного падежа вопросительного местоимения *чѐ*.

2) Существование достаточно большого числа нестандартных форм деепричастий на *-а / -я* (типа *положа*) и на *-ши* (типа *положивши*); к ним добавляются заимствованные из диалектов, но получающие всё большее распространение варианты на *-миши* (типа *вытимши, соврамши, не спамши* и т. п.)⁷.

3) Стойкое сохранение такого явления, как склоняемые краткие формы прилагательных в атрибутивной функции (*всяк человек, кари очи, тѣмну силу*, и т. п.); распространенность этого явления (в современном языке в основном связанного с использованием псевдонародной речевой маски) выходит далеко за пределы клишированных сочетаний типа *среди бела дня*. В недавней работе [Кулёва 2008] показана особая типичность таких «усеченных прилагательных» для языка русской поэзии практически во все периоды ее существования, вплоть до современной поэзии, где сочетания типа *пластмассовы цветочки* в определенных типах текстов и у определенных авторов оказываются вполне обычны.

Список этот, бесспорно, может быть продолжен (подробнее см. также [Ляшевская и др. 2005]), но уже и сказанного, на наш взгляд, достаточно, чтобы сделать вывод: грамматика русского языка, основанная на корпусе, будет весьма сильно отличаться от аналогичных сочинений, созданных в «докорпусную» эпоху.

Л и т е р а т у р а

Гиро-Вебер 1996 — М. Г и р о - В е б е р. Бисинхронный метод описания прилагательного в предикативной позиции в современном русском языке // А. В. Бон-

⁷ Вот характерный пример употребления такой формы из современного публицистического текста: *Из дальнейшего текста послесловия делается ясно, кого Макс Фрай, сбросимши прежних идолов с корабля современности, водрузил на их место* (М. Бутов. Отчуждение славой // «Новый Мир», 2000).

дарко (ред.). Теория функциональной грамматики: Качественность. Количественность. СПб., 1996. С. 65—79.

Зализняк 2004 — А. А. З а л и з н я к. Древненовгородский диалект. 2-е изд. М., 2004.

Кулева 2008 — А. С. К у л е в а. Усеченные прилагательные в языке русской поэзии: Дис. ... канд. филол. наук. М., 2008.

Ляшевская 2005 — О. Н. Л я ш е в с к а я и др. О морфологическом стандарте Национального корпуса русского языка // Национальный корпус русского языка: 2003—2005. М., 2005. С. 111—135.

НКРЯ 2005 — Национальный корпус русского языка: 2003—2005. М., 2005.

Перцов 2006а — Н. В. П е р ц о в. О роли корпусов в лингвистических исследованиях // Труды международной конференции «Корпусная лингвистика—2006». СПб., 2006. С. 318—331.

Перцов 2006б — Н. В. П е р ц о в. К суждениям о фактах русского языка в свете корпусных данных // Рус. яз. в науч. освещении. 2006б. № 1(11). С. 227—245.

Резникова 2008 — Т. И. Р е з н и к о в а. Корпуса славянских языков в интернете: Обзор ресурсов // Die Welt der Slaven, 2008. LIII. С. 10—38.

Резникова, Копотев 2005 — Т. И. Р е з н и к о в а, М. В. К о п о т е в. Лингвистически аннотированные корпуса русского языка (обзор общедоступных ресурсов) // Национальный корпус русского языка: 2003—2005. М., 2005. С. 31—61.

Тестелец 2001 — Я. Г. Т е с т е л е ц. Введение в общий синтаксис. М., 2001.

Трубецкой 2004 — Письма и заметки Н. С. Т р у б е ц к о г о / Подгот. к изд. Р. Якобсоном. М., 2004.

Bybee 2001 — J. B y b e e. Phonology and Language Use. Cambridge, 2001.

Bybee 2006 — J. B y b e e. Frequency of Use and the Organization of Language. Oxford, 2006.

Dąbrowska 1997 — E. D a b r o w s k a. The LAD goes to school: a cautionary tale for nativists // Linguistics. 35/4. 1997. P. 735—766.

Davies, Dubinsky 2004 — W. D. D a v i e s, S. D u b i n s k y. The Grammar of Raising and Control: A course in syntactic argumentation. Oxford, 2004.

Garde 1988 — P. G a r d e. Pour une méthode bisynchrone // Travaux du Cercle linguistique d'Aix-en-Provence. 1988. 6. P. 63—78. (См. также: P. G a r d e. Le mot, l'accent, la phrase: Etudes de linguistique slave et générale. P.: Institut d'études slaves, 2006. P. 437—445.)

Greenberg 1979 — J. H. G r e e n b e r g. Rethinking linguistics diachronically // Language. 1979. 55. P. 275—290.

Postal 1974 — P. M. P o s t a l. On Raising: An inquiry into one rule of English grammar and its theoretical implications. Cambridge (MA), 1974.