

ЕРШОВ А. П.

МАШИННЫЙ ФОНД РУССКОГО ЯЗЫКА

(Внешняя постановка вопроса)

Между экономикой и языком как системами существует немало сходных черт [ср. 1]. И та и другая в миллиардах единичных актов человеческой деятельности обнаруживают одновременно как изменчивость, так и устойчивость. Объективные законы каждой из этих систем складываются из интегрирования индивидуальных актов сознания, которые тем не менее подчиняются общим закономерностям системы. Было бы очень важно, чтобы в лингвистике, как и в экономике, произошел переход от методов наблюдения к методам измерения. Ведь только измерительный подход в экономике позволил создать экономические модели, выдвинуть экономические методы управления и приблизить экономику к разряду точных наук. Конечно, этот переход давно наметился, но появление вычислительной техники дает этой проблеме совершенно новую постановку. Аналогия с экономикой показывает лингвистике как диапазон возможностей, так и величину дистанции, которую надо преодолеть, чтобы получать от языка прямые ответы на огромное количество нерешенных вопросов. В этом плане весьма продуктивным было бы осуществление диалога человека с ЭВМ на естественном языке при условии ограничения этого диалога так называемой «деловой прозой». Надежда на продуктивность этой гипотезы основана на другой гипотезе — о внутренней формализованности любого устойчивого процесса производства и возникающих в нем человеческих отношений. Любой прогресс в области построения моделей и алгоритмов останется, однако, академическим упражнением, если не будет решена важнейшая задача — создание **Машинного фонда русского языка**. Это — фундаментальная проблема, решение которой будет иметь очень большую научную, общекультурную и прикладную ценность и окажет огромное влияние на саму лингвистику и ее место в обществе. По существу должна произойти некоторая смена парадигмы этой науки. Создание **Машинного фонда языка**, безусловно, станет могущественным средством нашего знания о языке, знания полного, не отстающего от хода времени, детального и обобщенного вместе, подвижного и инерционного, накапливающего все предыдущие временные срезы. При этом **Машинный фонд языка** должен, по крайней мере, содержать полный словарь и генератор словоформ, а также формализованный словарь (тезаурус) языка.

Еще один аспект отношения «ЭВМ в языке» — это проблема овладения языком. Прежде всего следует отметить явное увеличение в настоящее время роли печатного, услышанного и произнесенного слова. Круг словесного общения сейчас очень интенсивен. Средства массовой информации многократно повышают интенсивность такого общения, хотя, возможно, и делают его не столь прямым. Феноменальный рост личных книжных собраний дополняет эту интенсификацию. Есть ряд свидетельств тому, что широкое внедрение учебных ЭВМ в школе позволит радикально изменить ситуацию с изучением русского языка. Первые опыты применения ЭВМ в школе показывают, что машина может стать активным партнером в учебном процессе, создавая, где нужно, игровую обстановку, увеличивая элемент состязательности, обеспечивая индивидуальность и мгновенность реакции, побуждая к сознательному знанию и одновременно стимулируя и тренируя всяческую моторику, закрепляющую добытое знание.

Другой аспект интенсификации работы с русским языком я бы назвал проблемой экспансии языка, имея в виду его роль как языка межнацио-

нального общения и как мирового языка. И здесь я не вижу никаких причин для того, чтобы уклоняться от культурного соревнования с другими мировыми языками, и прежде всего с английским. Большое значение приобретает в этой связи создание различных словарей, в том числе энциклопедических. Конечно, мы делаем в этом плане очень много. У нас есть замечательные словари. Например, вышел великолепный однотомный «Советский энциклопедический словарь». Но ведь его нужно издавать каждый год, поддерживая конъюнктурные изменения. Праздником нашей культуры стало появление Лермонтовской энциклопедии, но где энциклопедии Пушкина, Достоевского, Толстого, Чехова, Горького? Почему Академический словарь русского языка должен выходить один раз в пятнадцать-двадцать лет? Можно этому найти много причин и обнаружить их объективность. Вопрос, однако, состоит в том, как можно преодолеть обстоятельства, этими причинами вызванные.

Нельзя забывать, что словарное и, вообще, лексикографическое дело для ведущих европейских языков — английского, французского, немецкого — уже давно поставлено на машину и соединено с достижениями полиграфии, прежде всего с фото- и электронным набором. Можно и должно дорожить накопленными фондами в их теперешнем виде, развивать сложившиеся методы работы. Мне эта задача представляется очень важной. Она важна не только для повышения культурного уровня нашего общества и не только для усиления интегративных процессов в нашей стране, но и для реализации функции русского языка как языка международного общения. А для того, чтобы понять, что тут все не так просто, достаточно просмотреть международные научные журналы, которые публикуют статьи как на русском, так и на английском языках, и посмотреть, как изменилось соотношение языков опубликованных текстов за последние двадцать лет.

Продолжая эту линию, я хотел бы поговорить о «языке в ЭВМ». Сначала небольшое философское отступление. За последние десятилетия начинает складываться представление о трех формах существования материи — вещества, отражающего постоянство материи, энергии, отражающей движение материи, и информации, отражающей структуру материи и сам процесс ее познания. Эти три формы дают определенную периодизацию развития человеческой цивилизации: освоение вещества, создание материального производства, овладение принципом единства материи (от глубины тысячелетий до начала XX в.); затем освоение энергии, создание энергетического изобилия, овладение принципом единства энергии и ее связи с веществом (от XVIII до XXI вв.) и, наконец, освоение информации, создание полной информационной доступности, овладение принципом единства процессов обработки информации в естественных и искусственных системах (от XX века в будущее). Через развитие средств электро-связи, вычислительных средств, автоматики и робототехники нам предстоит где-то в середине XXI века войти в период «сплошной информатизации», который в своем внешнем выражении будет характеризоваться следующими показателями: грандиозной сетью электронной связи с полу-миллиардом входов; полной автоматизацией техносферы с несколькими миллиардами встроенных микро-ЭВМ, двумя или тремя сотнями миллионов персональных ЭВМ и интеллектуальных терминалов, подключенных к сети связи; десятимиллионной иерархией универсальных ЭВМ, поддерживающих управление обществом и сетью связи; переносом на машинные носители практически всей информации, циркулирующей в обществе.

Все это должно не только существовать, но и воплощать полноту и доступность знания, хранить в себе достоверную информационную модель мира, быть в постоянном употреблении всеми людьми, реализуя на новой основе принцип единства человеческого рода. Возникает вопрос: как человек будет общаться с этой «инфосферой», как он будет побуждать машины к действию, как он будет черпать из этого грандиозного фонда знания, как он будет относиться к новому жителю своего дома — компьютеру? При первых же размышлениях над этим вопросом сразу возникает уже достаточно хорошо известная номенклатура конкретных технических

проблем: организация диалога с базами данных; составление баз знаний и общение с ними (база знаний отличается от базы данных, грубо говоря, тем, что представляет теорию более высокого порядка: если в базе данных — конкретные факты, то в базе знаний — общие суждения и определенные дедуктивные свойства); извлечение смысла из текста в виде команд, фактов, энциклопедических знаний и т. п.; пополнение знаний (очень сложная задача, т. к. речь идет не о механическом добавлении нового факта или суждения, а об установлении всех возможных связей с уже накопленной системой знаний); машинный перевод; синтез текста. О последней проблеме я хотел бы поговорить несколько подробнее. Каждый из нас сейчас читает много текстов, которые почти исключительно написаны людьми. Мы должны готовиться к тому, что это человеческое начало в текстовой информации начнет уступать заметное место машине. Хочу предупредить, что я не имею в виду машинного стихосложения или любовных писем, составленных роботом. Я говорю о реальных вещах, связанных прежде всего с деловой прозой. Сейчас, например, автоматизация проектирования становится стержнем технического прогресса. Но неотъемлемой частью автоматического проектирования становится синтез технических описаний. Причем это не отдельные фразы или перечни слов в каких-то фиксированных таблицах. Это нормальный связный текст. Уклониться от рассмотрения такой проблемы нет никакой возможности. Во всем большем количестве конструкторских бюро изготовление технической документации — узкое место. Профессия «технического писателя» — самая дефицитная. Нет людей, которые имеют вкус к такой работе. Устранить человеческий язык из технической документации никак нельзя.

Кроме технической документации к этой синтетической прозе нужно будет отнести большое количество сводок, отчетов и других текстовых свидетельств. Здесь — мой особый призыв к лингвистам. Очень не хотелось бы обрекать деловую прозу на посредственность. На первый взгляд деловая проза бесконечно далека от интереснейших задач внутреннего изучения русского языка, например, от изучения поэтических рядов. Есть ли у деловой прозы коммуникационная сверхфункция, есть ли у нее одухотворенность и чем она может быть красива — все это тоже, как мне кажется, достойнейшая задача науки о русском языке.

Я позволю себе закончить еще одной аналогией. Надеюсь, что она подчеркнет объем и размах предстоящей работы.

Мы хотим как можно глубже познать природу языка и, в частности, русского языка. Одним из выражений этого познания должна стать модель русского языка. Это формальная система, которая должна быть адекватной и равнообъемной живому организму языка, но в то же время она должна быть анатомически препарированной, доступной для наблюдения, изучения и изменения. Я хочу сопоставить эту еще не существующую модель с другой моделью, созданной человечеством для изучения неживой природы, — математическим анализом и даже с одной лишь его ветвью — дифференциальными уравнениями. Нет никаких сомнений в том, что по сравнению с русским языком — это куда более скромная область. И в то же время интеллектуальный багаж этой части науки огромен. Сколько мы имеем разных руководств, теорий, сколько методов решений дифференциальных уравнений, как велик поток литературы, как тщательно мы учим этому в университетах и технических вузах. Если сравнить размах области знаний, связанных с моделированием форм неживой материи, с нынешним статусом и объемом науки о языке, то станет ясно, что мы только в самом начале пути. Думается, что лишь сознание необходимости и неотвратимости качественного перелома в науке о языке, сознание ее общенаучной и государственной важности смогут создать тот запал и энтузиазм, без которого проблему Машинного фонда русского языка решить невозможно.

ЛИТЕРАТУРА

1. Rossi-Landi F. Linguistics and economics. The Hague-Paris, 1975.
2. Ершов А. П. Методологические предпосылки продуктивного диалога с ЭВМ на естественном языке. — ВФ, 1981, № 8.