

Дмитрий Жуков

МЫ—ПЕРЕВОДЧИКИ

ИЗДАТЕЛЬСТВО
«ЗНАНИЕ»
Москва
1975

4
Ж86

Жуков Д. А.
Ж86 Мы — переводчики. М., «Знание», 1975.
112 с. (Прочти, товарищ!)

Автор книги в увлекательной и доступной для широкого читателя форме разъясняет существо машинного перевода, описывает историю вопроса и на примере одного из научных коллективов показывает процесс создания алгоритма машинного перевода текстов с английского языка на русский.

Ж $\frac{70100-143}{073 (02) - 74}$ 165-74

Дмитрий Анатольевич ЖУКОВ

МЫ — ПЕРЕВОДЧИКИ

Редактор В. М. Климачева
Художник В. И. Пантелеев
Худож. редактор В. Н. Конюхов
Техн. редактор И. Г. Федотова
Корректор И. Л. Казеко

А 10810. Индекс заказа 47708. Сдано в набор 14/VI 1974 г. Подписано к печати 8/XII 1974 г. Формат бумаги 70X108¹/₁₆. Бумага типографская №1. Бум. л. 1,75. Печ. л. 3,5. Усл.-печ. л. 4,9. Уч.-изд. л. 4,94. Тираж 50.000 экз. Издательство «Знание». 101835 Москва, Центр, проезд Серова, д. 3/4. Заказ 1377. Цена 16 коп.

Киевская книжная фабрика республиканского производственного объединения «Полиграфкнига» Госкомиздата УССР, ул. Воровского. 24.

Издательство «Знание», 1975 г.

ПРЕДИСЛОВИЕ

К

нига Д. Жукова посвящена проблеме автоматического (машинного) перевода текстов с одного языка на другой. Эта проблема возникла более двух десятков лет назад. Ее существо сводится к построению автомата, на вход которого поступает текст на одном естественном языке (например, английском), а на выходе порождается текст на другом языке (например, русском). Сейчас в роли упомянутого автомата используются электронные вычислительные машины, для которых разрабатываются алгоритмы перевода и словари.

Современный этап развития машинного перевода может быть охарактеризован как этап синтаксического пословного перевода. В качестве основной единицы смысла здесь выступает слово, а грамматические формы и порядок следования слов в порождаемом тексте определяются на основе синтаксических связей между словами в тексте-оригинале. По аналогии со сложившимися традициями словари для автоматического перевода строятся преимущественно как словари слов (доля словосочетаний в них невелика).

Между тем в естественных языках слово не является единственной единицей смысла. В них одновременно используется несколько уровней семантических единиц (уровень морфем, уровень слов, уровень словосочетаний, уровень предложений и др.), причем смысл единиц более высокого уровня далеко не всегда может быть определен исходя из смысла единиц более низкого уровня (наиболее ярко это проявляется в отношениях между уровнями морфем и слов). Следовательно, и словари для автоматического перевода должны содержать единицы различных уровней.

Многовековой быт общения народов, говорящих и пишущих на различных языках, свидетельствует о том, что хороший переводчик должен не только понимать смысл переводимого текста, но и владеть фразеологическим богатством языка, на который осуществляется перевод. Таким образом, ручной перевод текстов представляет собой скорее фразеологический, чем пословный. Это обстоятельство нельзя не учитывать при автоматическом переводе.

На наш взгляд, качественный машинный перевод научно технических и деловых текстов (о художественных пока что говорить не приходится) может быть построен на основе использования комплексов словарей, включающих в свой состав семантические единицы различных уровней слова, словосочетания, предложения, высказывательные формы (устойчивые фразеологические обороты с пустыми местами, которые могут заполняться различными словами и словосочетаниями) При этом ведущее место должны занимать словосочетания, так как в указанных текстах наименования понятий чаще всего выражаются словосочетаниями и значительно реже отдельными словами В процессе перевода должно соблюдаться правило предпочтения, согласно которому прежде всего следует вычлениать и переводить семантические единицы высших уровней, а к единицам более низких уровней обращаться во вторую очередь и только в том случае, если нет иной возможности выполнить перевод

Важное место в системах машинного перевода должна занимать автоматизированная словарная служба, в задачу которой входит создание и ведение (дополнение, корректировка) словарей Для одной тематической области комплекс словарей может иметь объем 10^5 — 10^6 элементов (фраз, высказывательных форм, словосочетаний и отдельных слов), а количество тематических областей может исчисляться многими десятками Роль человека в автоматизированной словарной службе будет заключаться в разметке иностранных текстов с целью выделения из них элементов словарей и назначении последним соответствующих эквивалентов на выходном языке Разметка текстов может быть переложена и на машину

В настоящее время проблема машинного перевода в полном объеме еще не решена. Для ее решения необходимы не только глубокие теоретические исследования, но и большая по объему экспериментальная работа Нужно создавать крупные, хорошо организованные научные коллективы, нужна современная мощная электронная вычислительная техника с достаточным математическим (программным) обеспечением ее работы

Исследования в области машинного перевода следует отнести к разряду фундаментальных исследований Их нужно вести не только и не столько потому, что это позволит в будущем сэкономить средства на переводах иностранных текстов и облегчить процесс обмена научно технической и деловой информацией между различными странами Они необходимы как один из наиболее эффективных путей поиска технических решений по ряду проблем автоматической обработки информации, а уровень развития средств и методов автоматической обработки информации оказывает непосредственное влияние на темпы научно-технического прогресса

Г. Белоногов, доктор технических наук

ПЕРЕВОДЧИКИ В ЛАВИНЕ ИНФОРМАЦИИ

Н

а сколько языках говорят жители Земли? Количества их до сих пор не определил ни один лингвист. Предполагается, что даже число шесть тысяч не будет преувеличением. Еще в древние времена люди удивлялись великому множеству и разнообразию языков и наречий. Проходили века — племена кочевали, языки сливались и изменялись, но по-прежнему их было много, и это мешало общению между народами.

Ныне почти все народы известны, большинство языков описано, и сам мир стал удивительно мал благодаря великолепным средствам передвижения и связи. Но по-прежнему языковые различия препятствуют тесному общению и обмену достижениями между народами.

Мир по-прежнему говорит на тысячах языков, и становится понятной взволнованность известного английского ученого Джона Бернала, который писал, что «нужно выработать радикально лучшее средство общения, особенно ныне, когда мир становится действующим научным и экономическим комплексом, в котором вавилонская мешанина языков является ужасающими путями. Поистине дикое зрелище представляет собой множество людей, собравшихся на научную конференцию, которые совершенно одинаково одеты, одинаково выглядят, охватывают, как нам отлично известно, своими мыслями и познаниями почти тождественные области

знания и все-таки абсолютно не способны общаться между собой и нуждаются в услугах переводчика, если они не потратили уйму труда на изучение трех или четырех иностранных языков».

Одержимый упрямым неверием в возможность того, что разные народы найдут когда-нибудь общий язык, английский поэт Киплинг в свое время написал:

О, Запад есть Запад, Восток есть Восток,
и с мест они не сойдут,
Пока не явятся небо с землей на страшный
господень суд.

Лучшие умы мира задумывались над тем, как ликвидировать языковые барьеры между народами. Одни искали выход в создании единого искусственного языка для всех живущих на земле людей. Разработкой такого языка занимались крупнейшие ученые и писатели — от Декарта до Гюго, от Гумбольдта до Рассела. Другие старались упростить существующие языки. Так, в «бейсик инглиш» (упрощенном английском) всего 850 слов. Иные черпали грамматические правила и лексику из разных языков.

Изобретение каждого нового искусственного языка сопровождалось великой помпой. Изобретатели этих языков провозглашали, что найдена панацея от всех бед человеческих. Они основывали настоящие международные ордена, члены которых обязывались говорить только на новом языке. Эти ордена насчитывали тысячи приверженцев и пользовались покровительством римских пап и прочих духовных и светских владык.

Но народы не принимали искусственных языков. Они не собирались отказываться от своих культурных традиций, воплощенных в неимоверно богатых возможностях национального языка.

По данным ЮНЕСКО, сейчас во всем мире искусственным языком «эсперанто» пытаются заниматься восемь

миллионов человек. А всего за первую половину XX века создано свыше четырехсот искусственных языков.

Можно отдать должное упорству создателей искусственных языков и их сторонников, но нельзя не видеть, что все эти попытки обречены на неудачу.



Перевод с одного языка на другой — ремесло, существовавшее с незапамятных времен. 50—100 тысяч лет назад племена, говорившие на разных языках, общались друг с другом, и тогда уже была нужда в переводчиках.

Сначала были переводчики-любители, потом появились профессионалы.

Сначала переводили, не мудрствуя лукаво, потом появились теории перевода.

С расцветом художественной литературы заговорили о переводе как об искусстве.

Как переводить? Спор об этом ведется не первое тысячелетие. Одни требовали дословного перевода текста в ущерб языку, на который переводили. Так переводили Библию и трактаты Аристотеля. Так работал над Ветхим заветом святой Иероним, которого католическая церковь определила в покровители переводчиков.

Другие стремились донести до читателя лишь смысл текста. «Я полагал, что читатель будет требовать от меня точности не по счету, а — если можно так выразиться — по весу», — писал Цицерон о своих переводах речей Демосфена.

Много позже Дидро вообще пренебрег подлинником. Он прочитал книгу два раза, проникся ее духом, потом закрыл и стал переводить.

Сервантес не верил в силу перевода и вложил в уста Дон-Кихота скептическое сравнение его с изнанкой ковра.

Французские переводчики еще в XVII веке написали немало статей в защиту и буквального и вольного перевода. Со свойственной французам игривостью один из них сравнивал перевод с женщиной и говорил, что от перевода, как и от женщины, нельзя требовать, чтобы он был одновременно и красивым и верным.

Неверность перевода частенько бывала причиной дипломатических конфликтов. В Италии даже была поговорка: «Traduttori-traditori» — «Переводчики-предатели».

Вопросы перевода немало занимали и царя Петра I, при котором издавалось множество иностранных книг по военному делу, а также по наукам и технике. Он явно был противником буквализма и в своем «Указе Зотову об избегании в будущем ошибок» писал: «Г-н Зотов. Книгу о фортификации, которую вы переводили, мы оною прочли, и разговоры зело хорошо и внятно переведены, но как учит оной фортификацию делат... то зело темно и непонятно переведено.., И того ради надлежит вам и в той книжке, которую ныне переводите, остеретца в том, дабы внятнее перевесть, а особливо те места, который учат как делат; и не надлежит речь от речи хранить в переводе, но точию, сенс выразумев, на своем языке уже так писат, как внятнее может быт».

Можно привести тысячи высказываний о том, как надо переводить, и, пожалуй, мудрее всех и проще сказал Белинский, который не был переводчиком, но советовал переводить текст на русский язык «так, как бы написал его по-русски сам автор, если бы он был русским».

Когда-то считали перевод занятием, требующим больших специальных знаний, но малопочтенным. Философ Монтескье в своих «Персидских письмах» отказался считать перевод работой творческой.

Ныне неистощимое усердие переводчиков получило всеобщее признание. Перевод, и как искусство, и как ремесло, необыкновенно усовершенствовался.

И уже нет такой области, в которой бы не требовались услуги переводчиков.

Каждый год во всем мире переводится несколько десятков тысяч художественных книг. Но это ничтожная часть того, что переводится вообще.

Переводятся миллионы статей, патентов и инструкций...

Ежегодно проводится свыше тысячи международных конгрессов, симпозиумов, семинаров... и скромные труженики-переводчики помогают общению ученых, людей искусства и политических деятелей.

Перевод оказался в центре человеческих проблем. Жизнь требует тесного общения с другими народами, и рушатся древние понятия и предрассудки.



Это было в 1956 году.

Нас, в то время профессиональных переводчиков с иностранных языков, пригласили конструкторы электронных вычислительных машин, и задали нам вопрос, казалось бы, очень простой и имеющий прямое отношение к нашей профессии:

— Как вы переводите? Можете ли вы рассказать о процессе перевода во всех подробностях и по порядку?

Очевидно, наша группа переводчиков была не первой в том кабинете, куда привело нас приглашение, потому что в прищуре глаз спрашивавшего таилась каверза, которая стала ясна нам уже через несколько мгновений, прошедших после того, как смолк его голос.

Действительно, как мы переводим?

У одного из нас возникла в голове картина зала заседаний международного форума. Стеклопакетные будки переводчиков, обязанных переводить речи ораторов синхронно, то есть так, чтобы сидящие в зале слышали эти речи на родном языке. Ораторы не делают пауз для перевода, они говорят, говорят... Иные, увлекаясь, торопят-

ся, глотают окончания, а в наушниках, все-таки поспе-
вая, звучит скороговорка переводчика. Это очень тяже-
лый труд. Переводчики меняются каждые десять-пятнад-
цать минут. Частота их пульса достигает ста шестидеся-
ти ударов в минуту. По нагрузке на нервную систему
переводчиков сравнивают с летчиками, совершающими
полеты на сверхзвуковых скоростях.

Вот переводчик сидит с наушниками, в которых слы-
шится голос оратора, и говорит в микрофон на другом
языке, не слыша даже собственного голоса. Сперва, ког-
да он еще только учился переводить синхронно, его ох-
ватывал ужас, если он мешкал с переводом какой-ни-
будь фразы — оратор не ждал, уходил вперед на две,
три фразы... Что же сказал оратор? Слова как бы тая-
ли, смысл терялся, чувство ответственности сдавливало...
Но постепенно появился навык, память автоматически
удерживала в голове сразу несколько фраз, ум споро-
висто схватывал главное в каждой фразе, язык (опять
же автоматически) выговаривал стереотипные форму-
лировки, без которых редко обходится какой бы то ни
был оратор, на них переводчик как бы делает передыш-
ку и догоняет оратора... Так он переводит. Но об этом
ли его спросили?..

Другой представил себя дома за письменным сто-
лом. Перед ним раскрыт роман известного зарубежного
писателя. Сколько раз переводчик читал и перечитывал
его, перед тем как начать переводить! Он уже знал на
память почти весь текст романа, но все не решался
изложить его по-русски, так как в складывавшихся фра-
зах исчезало то трудноуловимое, что называется сти-
лем писателя. Оно — в интонациях, оно — в употребле-
нии определенных слов и языковых конструкций, оно —
в построении каждой фразы, оно — в том, как писатель
видит мир, в его образной системе... И если ты, пере-
водчик, даже влезешь в шкуру писателя, то все равно
написанное тобой на русском будет звучать не так,

и ты это знаешь лучше, чем кто-либо другой. Но ты продолжаешь искать, пока не находишь нечто подобное стилю писателя и на своем родном языке. А потом, когда начинаешь переводить фразу за фразой, тебя охватывает чувство радости — как гибок и прекрасен русский язык, какие возможности он открывает для передачи авторской речи, речи героев и персонажей!..

Впрочем, тут много подводных камней. Если ты переводишь с английского, то персонажи все-таки должны быть англичанами, а не русскими. Они не скажут: «По Сеньке и шапка», но выражение все равно должно быть народным. Проходят часы, а удачи все нет, и даже во сне придется эти персонажи и заговорят именно так, как тебе хотелось. Ты проснешься, запишешь, и дальше дело пойдет споро.

Иной раз думается, зачем томиться стремлением к совершенству каждой фразы, если сюжет романа интересен сам по себе? Сколько ты прочел переводных романов, язык которых безличен и гладок, как укатанная дорога? Будто переводили эти романы иностранцы, превосходно научившиеся русскому языку; у таких речь всегда гладка и как-то особенно грамматически правильна, но безжизненна. Она влияет на читателя, портит его речь, снимает ее национальную окраску. Не потому ли говорят, что язык, на который переводится художественное произведение, непременно должен быть родным языком переводчика? Но это лишь одна сторона работы. Чтобы понимать все тонкости иностранного языка, переводчик должен как можно лучше знать ту же, скажем, Англию, ее историю, культуру, народные обычаи, психический склад ее народа, природу, характерные особенности городских и сельских построек... Но его же спросили, как он переводит. Как ответить?

Третьему вспомнилось, как он после окончания института иностранных языков получил назначение в бю-

ро технической информации крупного завода и как при переводе первой же статьи из зарубежного технического журнала понял, что знание иностранного языка — лишь половина дела. Он не знал техники и потому никак не мог понять принципа действия прибора, который описывался в статье. Казалось бы, он знал значение всех слов (ему помогли технические и толковые словари, которые он горой взгромоздил на свой стол), но все равно общий смысл был темен. И тогда он взялся искоренять свою техническую неграмотность, сел за учебники, ходил по заводу, присматривался к работе механизмов и приборов, расспрашивал, и спустя некоторое время статьи перестали быть для него головоломками. Перед глазами вставала описываемая машина, он видел, как взаимодействуют ее детали, и переводы его стали весьма внятными для заводских инженеров. Как же он все-таки переводит? Ну, ищет незнакомые термины и слова в словаре, а потом... потом переводит. Как?..

Мы думали о трудностях, которые испытывали в своей работе, и это было психологически оправдано, но не отвечало на заданный нам вопрос. И вообще ответить на него оказалось непросто — мы еще вернемся к нему, но прежде узнаем, ради чего нас спрашивали, как мы переводим...



Путь познания сложен и бесконечен. Он извилист и напоминает отчасти дорогу в горах и одновременно лабиринт го множеством тупиков. Мысль выбирается из тенет укоренившихся представлений, делает рывок и, запутавшись в густых зарослях фактов, вновь упорно пробует себе путь на главную магистраль.

Разгадана тайна мыльных пузырей и созданы стройные гипотезы устройства Вселенной. Ничего, что завтра они могут быть опровергнуты и уступят место еще более хитроумным предположениям. В революционной

борьбе научных идей выковывается славное племя разведчиков незнаемого.

Человек видит, слышит, мыслит, говорит. О человеке думали, его изучали лучшие умы, и все-таки мы еще далеко не poznali самих себя. Как рождаются эти слова, которые возникают сейчас на бумаге? Что это за серое вещество головного мозга, в котором протекают удивительные процессы, преображающие ощущения в мысль?

Древнегреческий мудрец Аристотель утверждал, что мозг служит для выделения жидкости, охлаждающей сердце. И он же написал «Поэтику», в которой проследил тончайшие нюансы художественного творчества, выявил некоторые его законы, неопровержимые и в наши дни.

Странные вещи происходят с путями познания. Иной раз мы лучше знаем далекое, чем близкое. Описаны видимые миры и целые галактики, ракета уходит на миллионы километров по направлению к Марсу, а Земля, по которой мы ходим, остается неразведанной до конца. Мы проникли в глубь нашей планеты пока еще только на две тысячных ее радиуса.

Человеку было подвластно создание сложнейших абстракций, но до последнего времени он не имел почти никакого представления о работе собственного мозга. Но тем не менее человек мыслит и созидал.

Если не считать некоторых уродливых отклонений, всякое умственное усилие, исследование, обобщение, изобретение направлено на то, чтобы принести человечеству какие-то блага, защитить его от ярости слепых сил природы и заставить их служить, взять на себя часть бремени, которое несет человек, познавая и преобразуя природу.

Чем больше мы знаем, тем могущественнее мы становимся. Но с увеличением числа известных фактов процесс познания все усложняется.

Времена всеведающих энциклопедистов прошли давным-давно. Одна физика имеет сейчас сотни узких отраслей, каждая из которых накопила столько знаний, что их подчас трудно собрать воедино и систематизировать.

Существует афоризм, который гласит, что ученые знают все больше и больше во все меньшей и меньшей области знаний. Можно дойти до абсурдного предположения, что эта область станет в конце концов настолько микроскопически малой, что вообще будет нецелесообразно сию заниматься...

Если задуматься и отбросить шутки, то окажется, что тревога по этому поводу не лишена оснований.

Рецепта, как делать великие открытия, нет. Известно, что многие решения находили на стыках различных наук. Вот электронные вычислительные машины. Чтобы объяснить их работу, придется привлечь физику, математику, статистику, электронику и даже нейрофизиологию.

Более того, нередко методы какой-нибудь науки наталкивали на открытия в совершенно другой, отдаленной области знания. Так, языковед, подыскивающий формальный метод определения значений слов, вдруг наталкивается на решение в... ботаническом определителе растений.

Но попробуй уследи, что делается в других науках, если с каждым днем все труднее становится быть в курсе новшеств даже в своей, узкой области!

Все сведения, которые можно увидеть, прочесть или услышать, принято называть информацией. Обилие знаний породило мощный поток информации. И если объем знаний увеличивается вдвое каждые 50 лет, то объем информации удваивается каждые 10 лет, а за 50 лет он увеличится в 32 раза.

Одному английскому премьеру прошлого века приписывают определение трех степеней лжи. Он будто бы говорил, что «существует ложь, наглая ложь и статис-

тика». Но мы пишем не трактат об эпохе расцвета британской колониальной империи, и поэтому нам будет позволительно пренебречь политической иронией, заключенной в словах прожженного английского дипломата. Итак, обратимся к статистике.

В мире издается на 60 языках более 100 тысяч научно-технических журналов, в которых ежегодно публикуется до 5 миллионов статей. Каждый год регистрируются сотни тысяч патентов и пишется более четверти миллиона научных отчетов, диссертаций и т. д.

Ученые и инженеры тонут в этом необозримом море информации, пытаясь найти все нужное для работы.

Установлено, что отыскать статью или отчет о каком-нибудь экспериментальном исследовании подчас гораздо труднее, а следовательно, и дороже, чем провести эксперимент заново.

Инженеры шутят, что каждый день где-нибудь заново изобретается колесо. Это горькая шутка. Нередко из-за недостатка сведений создается телега, тогда как по улицам бегают автомобили.

Подсчитано, что неумение решить проблему информации стоит США не менее миллиарда долларов. А если подсчитать в масштабе всего человечества, то цифра будет поистине астрономической.

В нашей стране имеется несколько миллионов инженеров, более миллиона научных работников, и каждому из них нужны сведения о достижениях в их области знаний, так как отсутствие информации ведет к поискам уже давно найденных решений. Представьте себе, сколько средств тратится на бессмысленное дублирование и параллелизм?

Не пора ли прекратить нагнетание «ужасов», не пора ли, как на последней странице добропорядочного детективного романа, ткнуть перстом в кого-нибудь и сказать: «Вот он, виновник всех зол»?

Да, да, создается совершенно парадоксальное поло-

женис — нужные сведения трудно найти из-за обилия информации, а нехватка информации, в свою очередь, делает море сведений еще более обширным.

Ежегодно во всем мире выпускается 5 миллиардов экземпляров книг.

В библиотеке имени В. И. Ленина хранится около 30 миллионов книг, и ежегодно ее фонд увеличивается еще на миллион.

А наука не стоит на месте. Выходит все больше статей на русском, английском, французском, японском и прочих языках.

Происходит интернационализация науки. Крепнут международные связи. На глазах рушатся торговые барьеры. Наша страна заключает сотни и сотни крупных торговых сделок с фирмами капиталистических держав. На государственной основе происходит оживленный обмен научной и технической информацией. И естественно, растет необходимость в переводе научной и технической литературы, документов, патентов...

Только в Москве сосредоточены двенадцать мощных переводческих организаций. Всесоюзная торгово-промышленная палата имеет 18 отделов, занимающихся переводами. А сколько переводчиков трудится в отделах информации различных учреждений...

По неполным данным, в масштабе всей страны за один год переводятся с иностранных языков тексты, содержащие около трех миллиардов слов. Ежедневно в стране появляется 1000 разных переводов.

В 1972 году был создан Всесоюзный центр переводов научно-технической литературы и документации. В 1973 году он выпустил в печатном виде переводы, содержащие сто миллионов слов. В этом Центре 42 процента всех переводов осуществляется с английского языка, 25 — с немецкого, славянских и скандинавских языков, 10 — с французского и других романских языков, 13 — с японского и других восточных языков и 10% про-

центов — с русского на иностранные. Эти цифры весьма показательны.

Любопытно, что спрос на переводы Центром не удовлетворяется. Заказов у него вдвое больше.

Переводчики не успевают справляться с потоком зарубежной информации. Возможности человека ограничены. Опытный переводчик в среднем переводит восемь страниц машинописного текста за полный рабочий день.

Можно было бы еще увеличить число переводчиков. Но где их взять?

Пять лет учится будущий переводчик в институте, а потом ему еще требуется значительное время, чтобы приобрести хотя бы начатки специальных знаний и научиться понимать сложные технические и другие тексты.

А теперь представим себе работу современных конструкторов. Иные думают, что они создают новую машину или аппарат, конструируя его весь, до последнего винтика. Но это не так. Всякая новая конструкция состоит из бесчисленного количества решений, принятых прежде. Передатчик современного космического корабля состоит из уже известных деталей. А вот компоновка этих деталей, например, может придать ему новое качество.

Для миллионов инженеров информация нужна больше, чем для кого бы то ни было. Однако на сегодня в информации отражается не весь инженерный опыт.

Что же происходит? Неужели виноваты органы информации, из-за нерасторопности которых пропадают результаты труда и раздумий миллионов специалистов и ученых?

Не будем спешить с выводами. Вспомним, что объем информации удваивается каждые 10 лет. Если мы даже посадим читать, переводить и обрабатывать поступающую научную и техническую информацию еще 10 тысяч человек, то все равно будем напоминать чудака, пытающегося ложкой вычерпать реку.

●

Но вернемся в год тысяча девятьсот пятьдесят шестой. Уже тогда отмечали лавинообразный рост научной и технической информации. Но то же бурное развитие науки и техники сделало возможным создание быстродействующих электронных вычислительных машин. Вот тут-то и явилась на свет идея применить передовую технику для обработки информации, создать автоматические электронные системы, способные анализировать, накапливать и по первому требованию выдавать информацию. И даже переводить ее. Задача эта поистине исполинская, а по мнению иных, в то время и невыполнимая.

Технический прогресс становится невозможным без совершенной системы информации. И она будет создана. Более того, есть предположение, что в жизни людей такая система может занять место не меньшее, чем печатать. И одной из вех на этом пути должна стать переводческая машина.

Вопросы перевода являются делом государственной важности, в котором кровно заинтересованы миллионы специалистов и ученых.

Образованные и опытные переводчики обычно переводят до трехсот слов в час, включая время на поиски незнакомых слов в словаре и подготовку к работе. Но даже они иногда не могут уловить всех оттенков текста. К тому же человек быстро устает.

Существовавшие электронные вычислительные машины уже давали возможность осуществлять перевод со скоростью гораздо большей, чем триста слов в минуту. Но для того чтобы построить переводческую машину, надо было сначала классифицировать и связать идеи, научиться методам выражения этих идей, знать некоторые особенности человеческого мышления и логическую структуру языка. И все это изложить в виде рабочей про-

граммы, которой будет руководствоваться электронное устройство.

Вот с какой целью пригласили профессиональных переводчиков и задали вопрос, на который никто из нас не дал прямого ответа. Да его и быть не могло. Именно в те дни машинным переводом только начинали заниматься. Работа, которую нам предложили, и должна была стать ответом на сакраментальный вопрос.

Признаться, я и мои товарищи не сразу поверили в возможность автоматизации перевода. Мы долго рассуждали и спорили друг с другом.

— Послушайте, неужели это серьезно... Вы понимаете, что такое язык? Это не только сотни тысяч слов, это миллионы и даже миллиарды сочетаний, среди которых законными являются только те, что несут смысловую нагрузку. Математики, физики всегда имели дело с крупными числами. Вот в этом камешке, говорят они, миллионы молекул. Но ведь молекулы в нем похожи друг на друга, как две капли воды. А слова? Каждое полно значения и признает соседство лишь определенных собратьев...

Сейчас можно с улыбкой вспоминать свои сомнения, но тогда, в 1956 году, это приводило к серьезным треволнениям. Наше неверие порождалось незнанием. Никто из нас никогда и в глаза не видел электронной вычислительной машины. Нам было трудно поверить в то, что мы — именно мы, можем заставить какую-либо машину (и даже электронную и быстродействующую) делать переводы с одного языка на другой.

Весь наш переводческий опыт, все предубеждения, все наше существо восставало против такой возможности. Кому как не нам было знать, что перевод — занятие творческое... Переводчику необходимо уметь выходить из-под власти чужеземных языковых конструкций и писать чистую, красивую русскую прозу. Это бывает нередко труднее, чем просто писать на русском языке.

Неужели машина окажется способной овладеть красотами русской речи?..

Нам отвечали:

— Никто не покушается на красоты изящной словесности. Пусть она живет и здравствует во веки веков на радость переводчикам-виртуозам. На долю машины выпадет совсем другая работа... Несколько сот романов, пьес, рассказов может за год перевести сотня переводчиков, отлично знающих и чувствующих языки, знакомых с жизнью, которая описывается в художественных произведениях. Машинный же перевод необходим для снабжения современной технической и политической информацией тысяч учреждений, которым эта информация нужна, как воздух. Информационные тексты отличаются от художественных своей конкретностью и сравнительной простотой.

В то время в научных центрах многих городов Советского Союза уже велись работы по изучению строя русского, китайского, армянского, грузинского, венгерского, английского, французского и других языков. Но наша промышленность тогда еще не наладила серийного производства электронных вычислительных машин. Их было мало, и лишь немногие лингвисты имели возможность проверять свои выкладки на машинах.

Группы ученых были малочисленны, их разрозненные усилия не могли привести к практическому осуществлению машинного перевода.

К изучению языка привлекались статистические методы, математическая теория множеств и новейшая теория информации. Теоретические выкладки и подсчеты помогали раскрывать многие «тайны» языка, выявлять некоторые закономерности в построении фраз и сочетаниях слов. И действительно, эти закономерности оказались возможным представить в виде математических формул.

Но как было бы легко создать системы автоматиче-

ского перевода, если бы языки целиком и полностью подчинялись раз и навсегда установленным законам! Живая плоть языка, медленно, но верно меняющаяся, богатая великим многообразием форм, не желала укладываться в прокрустово ложе математических формул.

Мы посещали собрания лингвистов и математиков, где велись ожесточенные споры о будущем машинного перевода, выслушивали различные точки зрения, взвешивали их, оценивали, сравнивали. Приходя к себе, мы горячо обсуждали все услышанное. Надо было найти отправную точку для работы над машинным переводом с английского на русский, надо было освободиться от влияния ошибочных гипотез, которые могли бы увести в сторону и задержать создание универсальных правил перевода. Но для этого нам, как и читателю, требовалось узнать еще многое.

Нельзя сказать, что дело сразу пошло гладко. Многие из нас все еще сомневались в том, что машинный перевод возможен. Нам предстояло серьезно задуматься над «тайнами» своего ремесла. И попытаться записать логический процесс перевода пока хотя бы на бумаге.

Я вспоминаю поистине творческую атмосферу, царившую в лаборатории. Это было время учения и поисков, фантазирования и трезвых выкладок.

Каждый из нас был одновременно и преподавателем, и учеником. Разделив труд, мы тщательно готовились и читали друг другу лекции по основам формальной логики, теории перевода и лингвистике. Опытные программисты рассказывали нам, как осуществляется математическое программирование.

Сотрудники лаборатории машинного перевода знакомились с быстродействующими электронными вычислительными машинами, ибо, не зная их возможностей, лингвист-кибернетик в своих исканиях неизбежно обречен на неудачу. Как бы ни были остроумны его замыслы, они никогда не выйдут за рамки языковедения.

МАШИНА И ПАМЯТЬ



обываем в одном из вычислительных центров, где работают быстродействующие электронные вычислительные машины.

В длинном зале мало людей. У стены ферритовые кубы, под колпаками — быстро вращающиеся магнитные барабаны. Это «память» машины, хранящая нужные сведения и готовая в любую минуту послать их в арифметические устройства.

В центре зала — пульт управления. Под рукой у оператора десятки клавишей. На пульте несколько сот лампочек — индикаторов. Они то зажигаются, то гаснут, образуя сложные, ритмично повторяющиеся узоры. Каждая задача, которую решает машина, выводит свой световой рисунок. Игра света иногда дублируется звуками различной тональности, несущимися из динамика.

Опытное ухо оператора по этим звукам следит за ритмичностью работы машины. Шутники-математики вставили даже такие задачи-тесты, что, когда машина их решала, динамик начинал насвистывать «Барыню» или «Яблочко».

Машина производит сотни тысяч математических действий в секунду. Это уже не шутка. На такой машине можно рассчитать траекторию снаряда, прежде чем он долетит до цели, такие машины мгновенно определяют положение космического корабля в пространстве при малейших отклонениях от курса.

Нельзя не восхищаться этим совершенством, которое умно умом своих создателей.

Это на нее трудились физики в своих лабораториях, напряженно вглядываясь в зеленые пики на экранах осциллографов. Они создавали стандартные элементы, из которых потом сложилась машина. Автоматы тщательно проверяли качество деталей. Ученые думали о том, чтобы детали были надежны и служили как можно дольше. Ведь в электронной машине сотни тысяч элементов, и если они будут часто выходить из строя, машина потеряет право носить титул «быстродействующей».

Ученые позаботились о том, чтобы элементы машины не были громоздкими. Это они добились того, что микроскопические детальки ныне заменяют большие блоки с множеством электронных ламп.

А как создается такая машина?

...Она должна родиться через несколько лет, и в недрах конструкторского бюро начинаются горячие споры, какой ей быть.

Логики — инженеры и математики — под ритмичное постукивание портативных электрических арифмометров изучают свойства стандартных элементов, чертят диаграммы, стремясь к самым экономным, самым разумным решениям схем.

Иногда решение не приходит день, неделю, месяц, и логик, устав от раздумий, болтается по лабораториям и мешает работать другим. Вот он высказывает суждение о новом романе...

Он говорит, часто вскакивая, нервно бегаю по лаборатории, но взгляд его невнимательный, и чувствуется, что мысли его далеки от этой лихорадочной болтовни.

Потом он снова скрывается в кабинете, и в пять часов, когда всем пора уходить с работы, врывается к другим логикам и собирает их вокруг себя. Нашел! Ведь это так просто. Несложная логическая операция, и машинное время сокращено вдвое.

Куда там! Разве им докажешь! В новой идее множество прорех, но о том, что надо идти домой, забыли уже шестеро. Мысль их подстегнута. Они считают, пишут формулы, опровергают и от противного приходят к тому, что время можно сократить — хоть и не наполовину, но все же сократить. Машина будет производить в секунду еще на несколько тысяч операций больше.

Остроумные логические схемы обрастают значками триггеров и прочих элементов, вписываются в общую схему машины.

И вот уже конструкторы, склонившись у чертежных досок, увязывают стандартные элементы в компактные узлы. Они чертят стойки, шкаф, барабаны, пульт... Чертежи под слепящим светом копировальных аппаратов превращаются в розоватые «синьки», которые идут на заводы, изготавливающие узлы, и к монтажникам, собирающим машину.

Привезли и установили металлические стойки. Девушки-монтажницы не расстаются с паяльниками. Красные, желтые, синие проводники собираются в жгуты и змейками ползут от стойки к стойке, к пульту управления, к устройствам памяти.

Все собрано, смонтировано, подключена энергия, а машина, хоть убей, не работает или несет околесицу.

Начинается наладка, доводка. Постороннему человеку может показаться, что здесь царит неразбериха, но на самом деле идет планомерная, сосредоточенная работа. Споры неизбежны, в них — стремление прийти к истине, сделать все как можно лучше.

Бригада инженеров с тестерами в руках ползает буквально на коленях от контакта к контакту. Градом сыплются взаимные обвинения. Приходят логики и говорят, что, пока продолжалась вся эта возня, им в голову пришли новые «гениальные» мысли, а эта конструкция уже безнадежно устарела. От логиков отмахиваются, как от мух. Все считают, что машина может заработать только

чудом... И чудо случается. Машина работает, решает задачи.

Месяцы и месяцы работы, разочарований и удач позади. Но машина действительно «тихоходна», как черепаха. Подумаешь, сотня-другая тысяч логических действий в секунду. Дашь миллион! Пока машину опробовали, новые логические схемы уже перекочевали к конструкторам... В общем, конца этой истории не ждите, она бесконечна, как бесконечен технический прогресс.

Но вернемся в вычислительный центр. Эти геометрически правильно расположенные полупроводники, магнитные элементы, тонкие пленки, конденсаторы, пучки проводников носят печать какой-то своеобразной красоты. Красоты абстрактной и в то же время предельно конкретной для посвященных. Конструкторская эстетика заботится главным образом о компактности и удобстве для работы, и из этого стремления к целесообразности рождается новое понимание прекрасного, которое в наш технический век находит все больше и больше сторонников.

А что она умеет, эта машина? Вы внимательно взглядытесь в нее. Миллионы ячеек, из которых состоит наша глазная сетчатка, различают, выделяют отдельные детали. Сигналы, поступающие от наших органов чувств, интерпретируются миллиардами нейронов коры головного мозга.

Она может считать? Значит, это просто усложненная счетная линейка?

Вы снисходительно улыбаетесь, как будто вас знакомят с обезьяной, которую научили есть ножом и вилкой. Вы уже слышали, что вычислительная машина, которая может конкурировать с вашим мозгом, была бы величиной со стоэтажный дом.

Однако не торопитесь с выводами. Биотоки, которые генерирует ваш мозг, распространяются по нервным волокнам со скоростью нескольких десятков метров в се-

кунду. Электрический же ток распространяется по проводникам машины со скоростью 300 тысяч километров в секунду. Потому то машина и «соображает» намного быстрее человека. И главное, она может осуществлять простейшие логические операции, из которых складываются многие виды сложной умственной деятельности.

Задатки великолепные. Тем более, что есть даже некоторая аналогия с работой нашей нервной системы. Сведения, которые отправляются нашими органами чувств в мозг, и отдаваемые им распоряжения кодируются такими же условными сигналами, что и информация, которая вводится в машину. «Нейрограмма» состоит из сочетаний импульсов и пауз. И говорят, что это самый удобный код.

Импульс и пауза. Есть ток, нет тока. Положительный заряд, отрицательный заряд. «Да» и «нет». Единица и ноль.

Это самое меньшее, что мы можем узнать. Мы всегда стоим перед выбором из двух возможностей. Мы выбираем что-либо определенное, а потом в этом остальном выделяем опять что-нибудь определенное и опять, убедившись, что это «не то», переходим ко всему остальному, пока не добираться до истины... «Да» и «нет». Единица и ноль.

Мы подбрасываем монету и ждем, упадет она вверх загаданной стороной или нет. Такое сочетание из двух ответов принято считать мерой информации — мельчайшей крупницей знаний. Называют ее «бит» или «бинит» (сокращение от английского «байнери диджит» — двоичная единица).

Из сочетаний нулей и единиц состоит двоичная система исчисления, которой в наше время пользуются только дикие племена Торресова пролива и математики. Они утверждают, что такая система в некоторых отношениях даже удобней десятичной, которой привыкли пользоваться мы.

Числа 1, 2, 3, 4, 5, 6, 7, 12, 17 в двоичном исчислении выглядят так: 1, 10, 11, **100**, **101**, ПО, **111**, **1100**, **10001** и т. д. (число 2 будет читаться как «один — ноль»). Электронная машина способна складывать и вычитать эти числа, производить выбор (да, нет), сравнивать числа...

Для нас важно, что, обозначив, например, букву «а» числом «1», «б» — «2», «в» — «3» и т. д., мы можем вводить слова в машину, сравнивать их, находить нужные нам сочетания букв и слов в «памяти» машины и получать на свои вопросы односложные ответы: «да» или «нет».

Я не зря все время говорю о «да» и «нет». Это основа основ. Принятие любого решения начинается с выбора и кончается им. В классической логике есть так называемое дихотомическое деление — деление на два взаимоисключающих класса (**млекопитающие** и **немлекопитающие**, глаголы и неглаголы).

●

Для того чтобы машина работала, необходимо составить **алгоритм** — **систему** формальных правил, механическое выполнение которых приводит к бесспорному решению тех или иных задач. Сложный процесс может быть представлен в виде элементарных операций, которые выполняются электронной вычислительной машиной.

Элементарные операции осуществляются в арифметическом устройстве машины. Здесь электронные и магнитные приборы, на которые поступает электрический ток, «открываются» и «закрываются», намагничиваются и размагничиваются... Арифметическое устройство действует в точном соответствии с правилами сложения и вычитания двоичных чисел. Нам нет нужды подробно рассказывать о математической и технической сторонах этого вопроса, изложенных в десятках популярных книг по кибернетике.

Кроме арифметического устройства, в каждой вычис-

лительной электронной машине есть устройство ввода (оно преобразует числа, написанные или запечатленные в виде определенного расположения дырочек на перфокартах и перфолентах, в импульсы тока), запоминающее устройство (оно хранит введенные числа и результаты вычислений), устройство вывода (обычно буквопечатающий аппарат, преобразующий импульсы тока в информацию, понятную человеку) и устройство управления.

Подготавливая решение любой задачи с помощью электронной машины, математики-программисты составляют программу — последовательность команд, которые указывают машине, что ей делать, в какой последовательности осуществлять операции. Каждая команда говорит, в каком месте запоминающего устройства взять нужную информацию, каким логическим операциям ее подвергнуть и куда отправить результат вычислений.

Значит, решая проблему машинного перевода, нам прежде всего надо составить алгоритм, который затем математики преобразуют в систему команд для машины. Продумаем детально, что делаем мы, когда решаем математическую или лингвистическую задачу, разложим ход своих мыслей на простейшие логические операции и введем в машину.

И машина «запомнит» программу. Она «запомнит» и не «забудет» все сведения, которые ей придется обрабатывать.

«Память» у нее надежная и емкая.



Как-то мне довелось переводить (без помощи машины) рассказ известного американского писателя. Один из персонажей этого рассказа подслушивал разговоры и потом воспроизводил их публично с точностью магнитофона. Он передавал каждое слово, каждую интонацию. В том же рассказе есть упоминание и еще об одной исключительной памяти: слепой музыкант по имени Том

играл на рояле большие произведения, прослушав их в исполнении других музыкантов. Манера исполнения каждого пианиста, нюансы настроения и даже ошибки, которые они нарочно делали, — все это было в игре слепого музыканта. Феноменальная память обоих персонажей рассказа была бы достойна всяческого восхищения, если бы они не были... идиотами.

Забывать — это свойство человеческой памяти, которое еще ничего не говорит о степени умственных способностей.

Правда, из истории нам известны случаи исключительной памяти. Так, царь Митридат Понтийский знал 22 языка. Кардинал Меццофанти понимал 66 языков и мог говорить на 36. Александр Македонский и Юлий Цезарь знали имена всех своих солдат-соотечественников, то есть более 30 тысяч имен. Феноменальной способностью запоминать музыку обладал Моцарт, который в возрасте 14 лет после двукратного прослушивания в Сикстинской капелле четырех- и пятиголосных псалмов с десятиголосным заключительным хором записал их потом по памяти.

Великий шахматист Алехин давал сеансы одновременной игры на тридцати двух досках вслепую. Он свободно говорил на шести языках и, по утверждению Капабланки, «помнил наизусть все партии, иггранные мастерами в последние 15—20 лет».

Художник Н. Н. Ге собирался писать картину «Петр I допрашивает царевича Алексея Петровича в Петергофе». Ге посетил залу дворца, где некогда состоялся этот допрос, внимательно рассмотрел обстановку, вернулся к себе в мастерскую и написал картину, ни разу больше не наведавшись в Петергоф.

О замечательной памяти Наполеона рассказывает академик Тарле. «Однажды, посаженный за что-то на гауптвахту, он совершенно случайно нашел в помещении, где был заперт, неизвестно как попавший сюда

старый том юстиниановского сборника (по римскому праву). Он не только прочел его от доски до доски, но потом, почти 15 лет спустя, изумлял знаменитых французских юристов на заседаниях по выработке Наполеоновского кодекса, цитируя наизусть римские дигесты». Однако, зная склонность Наполеона к театральным эффектам, мы не можем поручиться, что он не перечитывал сборник перед заседаниями.

Можно было бы привести еще много исключительных случаев. Но способность обыкновенного человека к запоминанию значительно ниже. Если бы человек ничего не забывал, то все его силы тратились бы на запоминание непрерывно поступающей информации и это лишило бы человека возможности обобщать ее, мыслить, творить. Тогда человек перестал бы быть человеком и превратился даже, простите, не в идиота, а в подобие огромного рулона ленты для магнитной записи.

К счастью, этого не происходит. Человек забывает, но в то же время он нашел немало способов сделать так, чтобы информация, знания, сведения о материальном мире, в котором он живет, не исчезали бесследно.

Во-первых, он может для удобства классифицировать знания, раскладывать по полочкам и облегчать запоминание с помощью правил. Так, он умеет говорить на иностранных языках, пользуясь сравнительно небольшим числом грамматических правил.

Мы можем поместить в наш мозг в сотни раз больше знаний, чем у нас есть, ибо человек обладает способностью логически «свертывать» знания. (Кстати, этот же принцип применяется логиками для экономии ячеек памяти машины.)

Человек может улучшить свою память, применяя специальную тренировку ее. Еще в XIX веке немецкий психолог Эбингауз советовал заучивать списки из бессмысленных трехбуквенных слогов ради практики.

Люди хорошо запоминают только то, что постоянно

лежит в сфере их интересов. Иной раз вам трудно вспомнить, о чем говорилось в книге, которую вы прочли всего неделю назад, если эта книга по истории кино, а вы занимаетесь, например, резанием металлов. Зато с какой жадностью вы наброситесь на сообщение о технической новинке, связанной с вашей специальностью, прочтете его не раз, перескажете товарищам и, конечно, забудете не скоро. Знающие иностранные языки могут рассказать вам, как быстро забываются даже обиходные слова, если нет разговорной практики.

И вот тут можно сказать о втором способе сохранить информацию, которую не всегда в состоянии удерживать человеческая память. Неизвестно, с каких пор люди стали облегчать себе запоминание, делая зарубки на деревьях, вычерчивая и высекая на камне примитивные условные знаки. Потом они изобрели письменность, столь необходимую для того, чтобы запечатлеть знания, объем которых становился все больше и больше.

И это еще одно отличие современного человека от его предков, которые обладали не менее совершенным мыслительным аппаратом, но не имели в своем распоряжении колоссальной суммы знаний, накопленных за многие века ценой упорной работы, проделанной человечеством.

Наша дополнительная память — это и бесценные рукописи мыслителей, и книги, и киноленты.

И она огромна, эта память,— так огромна, что уже трудно выбирать из нее нужные сведения. Помочь тут могли бы быстродействующие электронные машины, которые будут «запоминать» неограниченное количество информации и выдавать ее по первому требованию. Они тоже будут нашей «дополнительной памятью».

«Дело заключается в том,— говорил академик Глушков,— что интеллектуальные способности человека в чисто количественном отношении столько же ограниче-

ны, как и его мускульная сила. Человек так же неспособен выучить наизусть все книги, хранящиеся в библиотеке, как и взвалить их себе на спину».



В свое время электронные машины были громоздки. Большие залы заполняли стойки с электроннолучевыми трубками, из которых строилась «память» машины. И работали такие машины сравнительно медленно, делая около 10 тысяч операций в секунду.

У большинства электронных быстродействующих машин, работающих сейчас в учреждениях и на предприятиях, запоминающие устройства еще далеко не совершенны. Большая часть сведений записана двоичным кодом, то есть сочетанием импульсов и пауз, на магнитных лентах. На рулоне ленты можно записать 80 миллионов битов информации. Но протянуть ее под считывающей головкой машины можно лишь за шесть минут.

Шесть минут — это очень большое время. Если нужные сведения находятся в середине ленты, то машине придется ждать их не менее трех минут. Эта «память» действует очень медленно и для современных машин служит как бы подсобным складом сведений, которые когда-нибудь потребуются ввести в машину.

Сейчас уже существует много видов машинной памяти. Например, магнитные барабаны.

Они действительно похожи на удлиненные китайские барабаны. Поверхность их покрыта магнитным слоем, как и у магнитофонной ленты. Запись двоичных знаков производится на магнитном слое сразу с нескольких сотен электромагнитных головок. Барабан вращается с огромной скоростью, и «считать» с него нужные сведения можно за ничтожную долю секунды. На каждом барабане записывается до полумиллиона двоичных знаков.

Еще большей емкостью обладает «память» на магнитных дисках с сотнями концентрических дорожек на каждом.

Ученые совершенствуют фотоскопическую память. Она представляет собой прозрачный диск, покрытый фотоэмульсией. На большом диске концентрически нанесены мельчайшие черные и белые прямоугольнички. Диск вращается со скоростью нескольких тысяч оборотов в минуту. Всего на нем более 1000 дорожек, и на каждой из них — десятки тысяч прямоугольников, чередование которых и является записью в двоичной системе.

Совершая один оборот за миллисекунды, диск прощупывается лучом электронной трубки, в которой возникают импульсы при прохождении через каждый светлый прямоугольник. Такая «память» запоминает до сотни миллионов знаков. Чтобы представить себе ее объем, достаточно сказать, что практически она может вместить в себя все слова русского языка.

Мы ознакомились с подвижными видами памяти. Мозг человека построен из механически неподвижных элементов. Из подобных же элементов ученые стараются строить и «память» машины.

Такова «память» на ферритах — колечках из пресованного порошка окислов железа и других металлов. Нанизанные тысячами на перекрещивающиеся тонкие проволочки, они способны «запоминать», в каком направлении течет ток, а следовательно, сведения, выраженные в двоичной системе. Их располагают стройными рядами в квадратной рамке величиной с почтовую карточку. Стопка рамок образует магнитный куб, в котором известен «адрес» каждого колечка. И в любое время легко найти в таком кубе необходимые сведения.

Сейчас десяток миллионов действий в секунду — не

предел для машины. И сама она претерпела значительные изменения в своем внешнем виде, сделалась более компактной и удобной в обращении.

Это стало возможным благодаря физикам, создающим все новые и новые стандартные элементы машин. Крошечные полупроводниковые транзисторные приборы пришли на смену громоздким электронным лампам.

Созданы целые блоки, которые собираются из микромодулей и весят всего несколько граммов.

В так называемых «криотронах» диаметр ячеек для хранения информации равен всего лишь одной стомиллионной доле сантиметра. А связывающие их проводники можно рассмотреть лишь в мощный микроскоп. Такие контуры и ячейки впечатываются в стекло и содержатся в жидком водороде или гелии при температуре, приближающейся к абсолютному нулю.

В этих условиях некоторые металлы почти не оказывают сопротивления электрическому току, и он течет без потери мощности.

«Память» машины все уменьшается и уменьшается по габаритам. Информацию стали записывать на тончайших магнитных пленках, которые слоями наносятся на схему методами напыления и осаждения.

Можно предвидеть создание сети информационных центров, которые будут автоматически собирать информацию, обобщать ее и по первому требованию сообщать по видеотелефону или фототелеграфу.

Уже сделаны первые шаги по пути создания таких центров.

Машины помогут удесятерить силы человека. Они не только возьмут на себя трудоемкую физическую работу, но и дадут возможность людям умственного труда «раскрепоститься» от наплыва фактов, которые нередко мешают сосредоточенному труду, и заняться «чистым» творчеством, пользуясь уже препарированными и обобщенными материалами.

НЕМНОГО ИСТОРИИ

И

дея создания машин, способных производить логические операции, существует очень давно. Джонатан Свифт более 200 лет тому назад в своих «Путешествиях Гулливера», высмеивая схоластические споры ученых того времени, писал, как в академии сказочной страны Лапуты один профессор изобрел станок, с помощью которого самый невежественный человек, произведя небольшие издержки и затратив немного физических усилий, якобы мог писать книги по философии, поэзии, праву, математике и богословию при полном отсутствии эрудиции и таланта. Станок выдавал сочетания слов, заранее написанных в различных наклонениях, падежах и временах. Иногда из хаотического смещения слов вдруг получалась осмысленная фраза.

Так в гротескной форме Свифт высмеял разговоры о логических машинах. Тогда они были бессмысленным прожектерством. Интересно другое. Значит, уже в XVIII веке между учеными велись споры, подобные нынешним дискуссиям.

Вообще гениальный сатирик Свифт оказался еще и гениальным провидцем. Вспомним спутники Марса или трех профессоров, предлагавших отказаться от употребления слов и заменить их вещами. Они таскали на себе громадные узлы с различными предметами и, показывая вещи друг другу, выражали таким образом свои мысли и желания. Посмеиваясь, Свифт говорит, что «великим

преимуществом этого изобретения является то, что им можно пользоваться как всемирным языком, понятным для всех цивилизованных наций, ибо мебель и домашняя утварь всюду одинакова или очень похожа, так что ее употребление легко может быть понято. Таким образом, посланники без труда могут говорить с иностранными королями и их министрами, язык которых им совершенно неизвестен».

При всей абсурдности «проекта» в нем заложена мысль об изобретении средства, которое облегчило бы общение между людьми, говорящими на разных языках. Им мог бы стать станок или устройство, способное логично манипулировать словами и «разбираться» в наклонениях, падежах и временах. Короче говоря, речь идет о механизации перевода.

Первый латино-французский словарь был составлен в 1498 году. В 1538 году некий Робер Этьен составил франко-латинский словарь. Словарь — первый элемент механизации перевода.

Долгое время машина для перевода с одного языка на другой мыслилась чем-то вроде механизированного словаря. Встретил переводчик незнакомое слово, нажал соответствующие кнопки и получил его эквивалент на другом языке. Но для людей, не знающих языка, она все равно не годилась бы. Трудно было бы догадаться, что английское выражение «Don't pull my leg», которое в буквальном переводе звучит так «Не тяни меня за ногу», на самом деле означает «Не морочь мне голову».

В тридцатые годы преподаватель истории науки и техники Петр Петрович Смирнов-Троянский сделал заявку на свое изобретение — «машину для автоматического производства нуждающихся только в литературной обработке готовых печатных переводов с одного языка одновременно на ряд других языков». Однако ему выдали авторское свидетельство за № 40995 на «машину для подбора и печатания слов при переводе с одного

языка на другой...», то есть на первый в истории техники механизированный словарь.

Это было механическое и с точки зрения нашей «электронной эры» очень медленно действующее устройство. Оно не получило распространения.

Но в автоматическом словаре П. П. Смирнова-Троянского уже была попытка решить некоторые лингвистические проблемы. Изобретатель исходил из того, что во многих языках порядок слов в предложении одинаков, и поэтому, если переводить слово за словом, смысл предложения можно понять. При каждом слове печатались его формы, а далее редактор должен был соединять набор слов в связный текст.

Если встречались слова омографы (то есть одинаковые по написанию, но разные по смыслу), то при них имелись такие указания:

Коса (песчаная)
Коса (девичья)
Коса (для косьбы)

Перевод (по службе)
Перевод (сочинения)
Перевод (стрелки)
Перевод (снятельной картинки).

В словарном поле машины Троянского помещалось 80 тысяч корней слов. Это много, если принять во внимание, что в русском языке 180 тысяч корней слов, а в английском — 200 тысяч^х.

¹ Созданный недавно в ФРГ автоматический электронный словарь служит большим подспорьем в работе переводчиков. Рабочие места переводчиков снабжены небольшими пультами с клавиатурой, через которые посылаются запросы в память электронной машины, хранящей переводы множества технических терминов. Число таких рабочих мест неограничено, и они могут находиться даже в других городах.

Самые благие идеи могут казаться неуместными и даже смешными, если нет средств для их осуществления. Но когда средства появляются, эти идеи начинают свою вторую жизнь. Они обретают сторонников, в них могут сомневаться люди осторожные, но зачеркнуть и высмеять их уже невозможно.

В 1944 году была создана первая быстродействующая электронная счетная машина, а уже в марте 1947 года профессор математики Массачусетского технологического института Норберт Винер получил письмо, в котором были такие строки:

«Мне хотелось бы задать вам один вопрос. Одной из серьезнейших проблем, стоящих перед ЮНЕСКО, проблемой, от которой зависят мир и созидание на нашей планете, является проблема перевода, так как она влияет на общение народов...

Полностью признавая (хотя и не вполне конкретно представляя себе) семантические трудности, возникающие вследствие многозначности слов и т. д., я подумал, нельзя ли сконструировать вычислительную машину, которая могла бы переводить. Даже если бы она переводила только научный материал (где семантических трудностей заметно меньше) и если бы получался неотработанный (но понятный) перевод, то, мне кажется, этим стоило бы заняться...

Даже если не знать ничего из официальных источников относительно новых мощных механических методов, применяемых в криптографии, а лишь строить о них догадки и умозаключения — методов, которые, мне кажется, успешно применяются даже тогда, когда не известно, какой язык зашифрован, — эти методы, естественно, заставляют задуматься над тем, нельзя ли рассматривать проблему перевода, как проблему криптографии. Когда я смотрю на статью, написанную по-русски, я

говору: «Это написано по-английски, но закодировано неизвестными символами. Сейчас начну расшифровывать».

Думали ли вы когда-нибудь об этом? Как лингвист и специалист по вычислительным машинам, считаете ли вы, что над этой проблемой надо подумать?».

Письмо было подписано членом Американской национальной академии наук Уорреном Уивером.

И Норберт Винер, уже работавший над своей книгой «Кибернетика, или управление и связь в животном и машине», тотчас ответил Уиверу:

«...Что касается проблемы механического перевода, то, откровенно говоря, я боюсь, что границы слов в разных языках слишком расплывчаты, а эмоциональные и интернациональные слова занимают слишком большое место в языке, чтобы какой-нибудь полумеханический способ перевода был многообещающим... В настоящее время механизация языка... представляется весьма преждевременной...»

Винер, знавший возможности вычислительных машин того времени, сомневался, что в их память можно ввести сколько-нибудь полный словарь, не говоря уже о точном переводе слов во всех их значениях.

Какие же причины побудили Уоррена Уивера обратиться с письмом к Винеру? Во время войны Уивер имел доступ к проектированию вычислительных машин, знал их возможности, их логическую гибкость. Знал он и то, что вычислительные машины широко применялись в криптографии.

На мысль о возможности применения машин для перевода его натолкнул рассказ одного известного математика П., который в свое время учился в Стамбульском университете и знал турецкий язык. К этому П. пришел друг, тоже математик, и сказал, что он разработал новый способ дешифровки. Он попросил П. составить зашифрованный текст, на котором можно было

бы испытать этот метод. П. написал по-турецки небольшой текст и упростил его, заменив турецкие буквы **ç, ğ, ı, õ, ş и ü на с, g, i, o, s и u**. Потом зашифровал текст, превратив его в колонку пятизначных чисел. Уже на следующий день друг принес ему текст, сказав при этом, что, очевидно, ничего не вышло. Но, добавил он, если получившийся ряд букв расчленить на слова и исправить некоторые буквы, то получится текст, похожий на турецкий.

Самое интересное, что друг П. не знал ни турецкого языка, ни того, что текст был написан по-турецки.

Постараемся теперь проследить дальнейший ход мыслей Уоррена **Уивера**. **Люди в разных** концах света создавали и развивали свои языки в общем при сходных условиях и сходным путем. Лексика (слова) и грамматика у всех языков разные, а вот логический строй, по-видимому, сходен, то есть мыслят люди, говорящие на разных языках, по одинаковому логическим законам. Это делает возможной дешифровку текстов людьми, — не знающими языка, но разбирающимися в логических закономерностях.

Все это убеждало **Уивера**, что формально машинный перевод осуществим. Сомневался он в одном: будет ли такой перевод литературным? «**Люди**, чувствующие все красоты русского языка, — писал Уивер, — говорят, что совершенно бесполезно пытаться перевести стихи Пушкина на какой-нибудь другой язык — этого не сможет сделать не только вычислительная машина, но и самый способный поэт, владеющий обоими языками».

И он был уже согласен на малое. «Не очаровывать или услаждать, не стремиться к красоте или изяществу, а приносить большую пользу в повседневной работе, делать доступным основное содержание документов (разрядка моя. — Д. Ж-), написанных на языках, не известных читателю».

Уоррену Уиверу в то время было трудно представить, как быстро разовьются «способности» электронных машин.

А пока он нашел единомышленника — доктора Эндрю Бута, который в отделении профессора Джона Бернала в Биркбекском колледже Лондонского университета занимался проектированием и конструированием вычислительных машин. В 1948 году Уивер посетил Бута в Лондоне и с восхищением следил за его первыми попытками изучить проблему создания автоматического словаря. Словарь этот давал еще только по одному значению всех слов подряд, что, конечно, не было переводом в нашем понимании этого слова.

В письме слово — это ряд букв, отделенных от других букв пробелами или знаками препинания. Так мы выделяем слово. Значит, по этим же признакам его может выделить и машина.

А как осуществляется перевод слова, как находится его эквивалент в другом языке? Предположим сначала, что слово однозначное, и представим себе упрощенно работу некоего абстрактного автоматического словаря.

Как поступаем мы? Нам кажется, что, встретив английское слово, мы просто отыскиваем в своей памяти его русский эквивалент.

Значит, если мы поместим в «память» машины словарь, то, сопоставляя встреченное слово поочередно со всеми словами английского словаря и получая ответы «нет», мы в конце концов дойдем до полного совпадения всех букв и получим ответ «да». Против каждого слова в словаре стоит какое-нибудь число, обозначающее место эквивалента этого слова в русском словаре. Теперь легко найти его и вставить в русский текст.

Только машинный словарь построен не совсем при-

вычно. Для того чтобы облегчить поиск нужного слова, он иногда строится сначала не по алфавиту, а по числу букв в каждом слове. В нем сгруппировались слова однобуквенные, двухбуквенные и т. д. А уже в каждой группе слова стоят по алфавиту.

Это делается для того, чтобы сократить время, которое затрачивает машина на поиск нужного слова. Ведь ей пришлось бы делать большое число сравнений, чтобы найти нужное слово.

Предположим, мы ввели в машину английское слово «and». Первый вопрос, который задается по программе, на русском языке прозвучал бы так: «В этом слове одна буква?» — «Нет», — ответила бы машина. «В этом слове две буквы?» — «Нет». — «В этом слове три буквы?» — «Да», — ответила бы машина. «Является ли первая буква слова буквой «а»?» — последовал бы вопрос. — «Да». — «Является ли вторая буква слова буквой «b»?» — «Нет»...

Вы уже, наверно, догадались, что в конце концов машина отыщет в своей памяти слово «and» и прочтет запечатленный рядом адрес русского эквивалента этого слова. По адресу в русском словаре легко отыщется союз «и». Другое дело, что «and» можно перевести как «а» и даже как «но». Это уже проблема многозначности...

Продельваем ли мы все эти операции, когда переводим сами? Когда мы еще только начинаем учиться, последовательность многих из этих действий заметна. Но потом совокупность признаков слова анализируется в нашем мозгу уже так быстро, что в нашей памяти всплывает сразу готовый эквивалент, а промежуточные операции как бы совсем исчезают.

Кстати, машина произведет все эти действия, описание которых заняло у нас столько места на бумаге, в тысячи раз быстрее, чем человек, вспоминающий знакомое слово. Не надо забывать, что машины производили уже миллион логических действий в секунду.

Но вернемся к Уиверу, которого мы оставили в лаборатории вычислительных машин Бута. Здесь он познакомился с работой доктора Риченса, который был помощником начальника Бюро выведения растений и генетики. Риченс много занимался вопросами реферирования материалов и так же, как Бут, заинтересовался проблемой машинного перевода.

Нетрудно заметить, что в нашем повествовании еще не встречались имена профессиональных лингвистов и переводчиков. Язык — достояние людей любой науки. Ученым приходится читать много информации на других языках и даже переводить. Из знания техники, логики и языка родилась мысль о создании машины-переводчика. Лингвисты же, круг интересов которых был ограничен только языковыми проблемами, просто не могли выйти за рамки своей науки.

Доктор Риченс, как и Бут, еще не занимался многозначностью и порядком слов, идиомами и т. д., но уже пришел к мысли, что машина может сама отчасти разбираться в грамматике.

Так, в автоматическом словаре Бута были записаны основы слов. Например, «бег». Но в тексте встретилось русское слово «бегущий». Такого слова в «памяти» машины нет. Она отбрасывает по одной букве с конца слова «бегущий» и сравнивает остаток со словами в **своей** «памяти». «Бегущий», «бегущ», «бегу», «бег». Такое **слово есть**. Тогда машина находит в грамматическом приложении к словарю остальную часть слова («ущий»). А там уже заранее стоят грамматические примечания (причастие, настоящее время, мужской род, единственное число, несовершенный вид).

В июле 1949 года Уоррен Уивер изложил свои соображения в меморандуме, который озаглавил «Перевод», и разослал его 200 знакомым ему специалистам в различных областях знаний. Это было провозглаше-

нием возможности перевода с одного языка на другой с помощью вычислительных устройств.

Одним меморандум Уивера показался откровением, другие насмешливо восклицали: «Чепуха!». Иные недоверчиво читали строки меморандума, но, поразмыслив, проникались его идеей и садились писать письма Уиверу, излагая все новые и новые соображения.

Пришел восторженный ответ от известного ученого Ванневара Буша. «Мне вовсе не кажется, что многозначность слов может оказаться препятствием для осуществления вашего проекта. Напротив, это обстоятельство я рассматриваю как фактор положительный. Разумеется, машина может сделать выбор одного из одновременно полученных ею сообщений. По-моему, вовсе нетрудно заставить переводящую машину выбрать нужное слово так же обоснованно, как это делают люди-переводчики, в особенности тогда, когда они переводят материалы по не знакомому им предмету. Другая сторона вопроса — порядок слов, и здесь, по моему мнению, тоже должно найтись нечто заслуживающее внимания. Машина легко может «держать в памяти» сделанный ею перевод предложения до тех пор, пока не дойдет до его конца, то есть до точки, после чего, как мне кажется, вполне возможно заставить ее выполнить некоторые операции на основании правил о порядке слов, прежде чем она выдаст окончательный перевод. Во всяком случае, как бы там ни было, я полагаю, что это дело может быть осуществлено самым блестящим образом».

И время показало, что восторженная оценка способностей машины, данная Бушем, оказалась реальной.

Меморандум Уивера пробудил интерес ученых к проблеме "перевода, и во многих университетах США и Англии началась исследовательская работа.

Уоррен Уивер, соавтор Клайда Шеннона по книге «Математическая теория связи», в которой излагались идеи этого знаменитого инженера, пробудив интерес специалистов к машинному переводу, больше к нему не обращался. Разносторонность этого ученого, его готовность поддержать любое начинание, идущее на благо людям, сказались и в его горячих и обоснованных выступлениях против генетической опасности, возникающей в результате ядерных испытаний.

Неудержимый прогресс советской науки и публикация большого количества ценных статей наших ученых привлекли внимание американцев к переводам с русского языка. Появилось много работ, посвященных структуре русской речи.

В Гарвардском университете был поставлен опыт. Студентам лаборатории численного анализа предложили разобраться в тексте работы академика Колмогорова. Механический словарь просто напечатал все английские варианты каждого русского слова. Статья начиналась словами «В последнее время...» И вместо первого же русского предлога «в» словарь напечатал шесть английских (in, at, into, to, for, on). Как говорится, на выбор дороже... Вот она проблема многозначности при переводе!

Студенты затратили много времени, приводя такой «перевод» в порядок. При этом надо учитывать, что обрабатывали текст специалисты, а уже давно доказано практикой, что опытнейший переводчик художественных произведений вряд ли сможет перевести доклад ученого-физика. Это будет легче сделать его коллеге — физику, если даже он обладает слабыми познаниями в языке.

Один из студентов написал в отчете: «В то время как значение... было сразу ясно, мне пришлось удивить»

тельно долго работать, чтобы перевести текст на литературный английский язык».

Очевидно было, что редактирование английского текста — не выход из положения. Машина сама должна переводить на правильный английский язык. Эту проблему и обсуждали 18 американских и английских ученых, съехавшихся весной 1952 года на первую конференцию по машинному переводу в Массачусетский технологический институт. Идея машинного перевода перестала казаться досужей выдумкой, ею увлекалось все больше лингвистов и создателей вычислительной техники.

И ДОГ 7 января 1954 года в нью-йоркской конторе фирмы «**Интернэшнл бизнес мэшинз**» (**ИБМ**) произошло событие, которое в истории машинного перевода получило название «Джорджтаунского эксперимента».

В Институте языка и лингвистики Джорджтаунского университета группа ученых во главе с Леоном Достертом подготовила для перевода с русского языка на английский небольшое число предложений, составила словарь из 250 слов и шесть правил перевода. Эти правила были запрограммированы Питером Шериданом из фирмы ИБМ, занимающейся производством электронных счетных машин.

В торжественной обстановке в машину ИБМ-701 на перфокартах вводились русские фразы: «Обработка повышает качество нефти», «Международное понимание является важным фактором в решении политических вопросов», — и машина выдавала английский перевод...

Надо сразу оговориться. Пышная демонстрация Джорджтаунского эксперимента прежде всего преследовала рекламные цели. Фирма **ИБМ** начала делать бизнес на всеобщем интересе публики к машинному переводу. Главное было — привлечь внимание к

товару фирмы, обладающему «уникальными способностями».

И тем не менее это был первый в мире эксперимент, во время которого машина перевела правильным английским языком 60 несложных русских предложений, состоявших только из слов, включенных в словарь. Некоторые русские слова уже имели по два английских значения. Например, в зависимости от конкретного следующего слова машина переводила русский предлог «к» английскими «to» или «for».

В словаре, помимо слов, были и некоторые русские падежные окончания. Падежей в английском языке нет, и поэтому в словаре содержалось указание, какие предлоги в английском переводе выполняют функции тех или иных русских падежей. Так, роль родительного падежа в английском языке часто играет предлог «of».

Ко времени Джорджтаунского эксперимента многие ученые уже определенно знали, что перевод — это не просто замена одних знаков другими. Иначе можно было бы сказать, что «переводит» стенографистка, записывая речь условными значками, или телеграфист переводит телеграмму на язык Морзе.

Неоправданным оказалось и предположение Уивера, будто проблему перевода надо рассматривать как проблему криптографии. Английская фраза — это не просто русская фраза, зашифрованная путем сложных **математических выкладок**. **Несмотря** на некоторое сходство логических структур языков, имеющиеся различия были настолько велики, что трудности, которые встали на пути машинного перевода, порой казались непреодолимыми.



1954 год. В одном из зданий на Ленинском проспекте в Москве уже работала быстродействующая счетная электронная машина (БЭСМ), фотографии которой впоследствии публиковались в журналах и газетах.

Она была сконструирована под руководством выдающегося ученого, академика Сергея Алексеевича Лебедева и установлена в левом крыле здания Института точной механики и вычислительной техники. Машина трудилась над решением очередной задачи с тысячами уравнений, требовавших сотен миллионов арифметических действий, высчитывала орбиты движения планет, определяла наивыгоднейшие конструкции мостов и реактивных двигателей и даже решала шахматные задачи.

Идея поставить первый опыт автоматического перевода на электронной машине в нашей стране впервые возникла в 1954 году в разговоре трех работников института — специалиста по электронным машинам И. С. Мухина и математиков Л. Н. Королева и С. Н. Разумовского. Они поделились своими соображениями с академиком С. А. Лебедевым и тогдашним директором ВИНТИ профессором Д. Ю. Пановым, которые поддержали начинание, а затем взяли на себя общее руководство этой работой.

Для работы над лингвистической частью проблемы Д. Ю. Панов рекомендовал кандидата филологических наук И. К. Вельскую. В январе 1955 года она приступила к работе, а уже в декабре того же года машина осуществила пробный перевод с английского отрывков из книги Милна «Численное решение дифференциальных уравнений». Ученые познакомились с работами американцев и решили идти своим путем.

Инженеры и математики объяснили Вельской основные принципы работы машины, возлагая большие надежды на ее обширные лингвистические познания, так как они понимали, что машинный перевод — не только техническая проблема, это прежде всего проблема языка.

Вельская прочла статью американца, доктора Эттингера, который, развивая взгляды Уивера, утверждал, что

процессы перевода и кодировки телеграфного сообщения сходны, что «в каждом из этих двух случаев процесс сводится к перекодированию сообщения с помощью новой системы символов».

Нет, Уивер был неправ. Перевод с одного языка на другой — это не частный случай криптографии.

Кроме общечеловеческих логических законов, у каждого национального языка есть свои логические законы, присущие только данному языку. Почему англичане говорят, что «лампа свисает с потолка», а мы — «лампа висит на потолке»? Почему мы говорим «на Кавказ», но «в Крым», а англичане и то и другое выражают одним предлогом направления «to»? И в то же время, почему англичане, переводя русскую фразу «я прибыл в город», передают один русский предлог «в» английскими «in» или «at» в зависимости от того, какой это город — большой или маленький.

Значит, думая на разных языках, мы руководствуемся разными ассоциациями и языковыми традициями, и общие логические законы помогают при переводе лишь отчасти. Надо брать каждый язык во всей его сложности.

Вельская со свойственной ей решительностью взялась за анализ каждого языка отдельно и стала искать возможности передавать формы английского языка средствами русского языка.

Язык сложен, но не хаотичен. Каждый язык представляет собой систему, имеющую свои определенные законы. В языке все значит, все имеет какую-то форму. «Поэтому,— писала впоследствии Вельская,— анализ оформления слова (в широком смысле) дает все необходимые сведения для правильного понимания слова в предложении.

Обычно у нас не возникает сомнения в том, что существует непосредственная связь между данной грамма-

тической формой слова и наличием или характерным отсутствием у него некоторого материального оформления. Но мысль о том, что и выбор значения многозначного слова может быть формально определен, кажется менее очевидной. Переводчики нередко склонны ссылаться на «интуицию», «общее содержание фразы» и тому подобные, «неформальные» причины, по которым они выбрали то или иное значение из нескольких возможных для данного слова.

Это характерное заблуждение вызвано тем, что закономерности языка, действующие в сфере лексики, много сложнее, чем грамматические законы. Поэтому их труднее систематизировать, увидеть общее за индивидуальным и сформулировать это общее как закон».

Бельская, разумеется, многое упрощала. Но иногда бывает полезно приуменьшить трудности. Хотя бы ради того, чтобы сдвинуть дело с мертвой точки...



Для первого опыта машинного перевода был составлен словарь из 952 английских и 1073 русских слов. Специальный математический текст (книга Милна) был заранее переведен на русский язык, но перевод этот отличался от обычного.

Читая переводные книги, мы видим лишь конечный результат усилий переводчиков. Мы не думаем об их затруднениях и заботах, об их ухищрениях и изобретательности.

Перевод английского текста, подготовленный для машины, не был простой цепью русских фраз. Читателю трудно было бы разобраться в многочисленных схемах и списках, составленных Вельской к концу 1955 года. Но именно эти схемы отражали сам процесс перевода, логику переводческой работы.

Этот процесс был разложен на простейшие логические операции. За проверкой на определенные признаки следовал ответ «да» или «нет», потом — новая проверка на новые признаки и так далее, пока не находился нужный ответ.

Д. Ю. Панов писал, что этот принцип очень похож на принцип определения растений или бабочек по специальным определителям, в которых указывается ряд признаков, подобранных так, что в зависимости от наличия или отсутствия их удается установить принадлежность растения или бабочки к тому или иному виду.

Так, в определителе может быть указан признак «пестиков несколько», но в цветке их нет. Тогда ученый отсылается к другим признакам. А если они есть, то дальше указывается точное число их и название растения.

Та же операция продлевается и со словами. Вот слово «example». Мы проверяем, нет ли перед ним слова «for». Если есть, то его можно перевести русским вводным словом «например». А если нет, то следует ряд других проверок и переводов, и, наконец, основное значение слова — «пример» (имя существительное, 2-го склонения, мужского рода).

И машина должна была слепо повторить за человеком эти операции. Математики с самого начала подробно объяснили Вельской возможности машины, а теперь они переводили схемы и правила, изложенные обычным русским языком, на язык чисел, понятный «электронному мозгу».

Программа перевода, составленная для узкоспециальной книги Милна, годилась и для других текстов, но для этого в нее добавлялись новые слова и правила. Однако это был первый удачный шаг на пути к решению проблемы перевода с одного языка на другой.

Вскоре в Институте прикладной математики Академии наук, на машине «Стрела» по инициативе члена-корреспондента Академии наук А. А. Ляпунова под

руководством О. С. Кулагиной был сделан еще один опытный перевод научного текста. На этот раз с французского на русский.

Подводя итоги первых экспериментов И. С. Мухин писал: «Разумеется, от первых опытов автоматического перевода, которые осуществлены в настоящее время в СССР и за рубежом, еще далеко до практической реализации автоматического перевода в сколько-нибудь крупных масштабах. Однако есть все основания ожидать в ближайшем будущем новых успехов, по крайней мере в деле перевода научно-технического текста».

Первые опыты обнадеживали. Но пора было подумать о создании такого алгоритма, с помощью которого машина могла бы осуществлять не опытный, а, если так можно выразиться, «промышленный» перевод. То есть она должна была стать рентабельной и переводить огромные массивы информации, заменяя сотни переводчиков, на подготовку которых затрачивается так много времени, труда и средств.



Итак, в область языкознания вторгалась техника. Профессор Московского университета В. А. Звегинцев так рассказывал об этом вторжении:

«Когда впервые на практике была доказана возможность машинного перевода, инженеры и математики-программисты пришли к нам, лингвистам, и потребовали: «Дайте нам строгие правила перевода для множества языков, дайте нам точные законы для перевода вообще, чтобы мы могли разрабатывать конструкции переводческих машин и программы для них. Науку захлестывает обилие фактов и исследований, изложенных на разных языках, скоро без машин тут обойтись будет нельзя»...

В готовом виде таких правил не существовало, но недостатка в лингвистах и математиках, желавших при-

ступить к изучению законов языка, у нас не было с самого дня зарождения машинного перевода.

На первых порах, когда идеи машинного перевода еще только получали право на существование, слышалось немало горячих речей и щедрых обещаний. Энтузиастам казалось, что машинный перевод — дело простое, что надо лишь взять богатый словесный материал, накопленный в словарях, добавить к нему сведения из грамматик, запрограммировать все это, и машина начнет переводить.

Но оказалось, что даже составление словарей для машинного перевода — дело очень сложное, требующее многолетней кропотливой работы и применения технических средств, а традиционные грамматики неточны в своих выводах, и использовать их для нужд машинного перевода можно только как подсобный материал.

Как и всякое новое дело, машинный перевод имел своих яростных противников и не менее яростных приверженцев. В лагере его приверженцев тоже не было единомыслия, как нет его и до сих пор. Правда, эта борьба не приобретала драматического накала, и сначала было бы смешно побивать друг друга умозрительными заключениями. Борьба выливалась в мирные дискуссии о развитии машинного перевода.

Впоследствии одни ученые занялись практическим осуществлением выдвинутых ими идей, другие по-прежнему дискутировали, считая себя теоретиками машинного перевода. Но сейчас уже накоплено достаточно опыта, чтобы оценить по достоинству различные точки зрения.

К концу пятидесятых годов разногласия эти не получили своего разрешения, хотя некоторые уже были твердо уверены, что стремление «к полностью автоматизированному производству высококачественных переводов» имеет под собой твердую почву. Тому порукой было головокружительно быстрое развитие электрон-

ной техники и автоматизации исследовательских работ при создании алгоритмов.

И технические предпосылки к созданию машинного перевода уже были в 1961 году, когда профессор Колумбийского университета Мортимер Таубе, выпустивший в свет свою едкую книгу «Вычислительные машины и здравый смысл» («Миф о думающих машинах»), предупреждал, что без научного обоснования инженерной реализации машинного перевода «в свете известной неформальности языка и смысла» изыскания в этой области «носят характер не истинно научных исследований, а романтического поиска Грааля».

В Советском Союзе вопросами прикладной лингвистики и автоматизации перевода занимались десятки лабораторий и групп, в которых работали сотни людей. В Москве проблемы машинного перевода изучали сотрудники Всесоюзного института научной и технической информации (ВИНИТИ), Центрального научно-исследовательского института патентной информации (ЦНИИПИ), Математического института имени В.А. Стеклова и других научных учреждений и учебных заведений. Несколько групп работало в Ленинграде, некоторые вопросы решались в Киеве, Новосибирске, Ереване, Тбилиси, Горьком и других городах.

В 1958 году была созвана первая в СССР конференция по машинному переводу. Большинство ее участников не соглашались с «эмпирическим подходом» к решению проблемы машинного перевода, то есть с принципами, которыми руководствовалась И. К. Вельская.

Некоторые из них считали, что делать «двуязычные» или «бинарные» алгоритмы невыгодно. Зачем заниматься переводом только с одного языка на другой, говорили они, кпгда можно создать такой алгоритм, который будет переводить с любого языка на любой другой язык? Если мы возьмем четыре языка и будем переводить с каждого на каждый, то понадобится двенадцать

двуязычных алгоритмов. Для десяти языков алгоритмов уже будет почти сто.

Но как же сделать такую «всепереводящую» машину? И предлагалось решение. Надо, мол, создать единый язык или, как его называют, язык-посредник. Машина будет сначала переводить с любого языка на язык-посредник, а потом на любой другой.

И снова разногласия. Одни предлагали взять за язык-посредник один из «живых» языков, скажем, английский или русский.

Другие утверждали, что «живые» языки очень сложны, многозначны, имеют много особенностей, присущих только им, и даже исключений из этих особенностей. Надо, говорили они, сделать языком-посредником какой-нибудь искусственный язык. Например, эсперанто. У него простая грамматика и нет исключений.

Нет, говорили третьи, такой язык был бы слишком примитивен. С «живого» языка на него ничего толком не переведешь. Слишком беден был бы такой перевод. Надо создавать нечто среднее между эсперанто и живым языком. Н. Д. Андреев, руководитель лаборатории машинного перевода при Ленинградском университете, предложил создать язык-посредник на основании грамматики и словарей различных «живых» языков и выразить его в символах, «понятных» машине. Андреев как бы «усреднял» реальные языки мира, включая в свой язык-посредник только самые типичные, самые частые грамматические правила и слова, общие для большинства человеческих языков.

Четвертые считали, что такой язык мало чем отличался бы от эсперанто, и предлагали создать язык-посредник, который бы содержал все категории всех языков. Он был бы богаче любого из живых языков.

Но никто из сторонников языка-посредника пока так и не добился ощутимых практических результатов.

Высказывался ряд мнений, что грамматику языков

надо упрощать, сжимать в объеме. Призыв упрощать родился из неверия в перспективу развития электронных машин. Поговаривали даже о том, что существующие обычные машины не годятся для перевода и надо создавать специализированную технику. Сначала это имело некоторый смысл. Машины еще не обладали большой памятью и быстродействием, и упрощение языка позволяло экономить машинное время. С совершенствованием машин появилась возможность предусматривать языковые тонкости, которые при математическом подходе отчасти теряются.

Многие принялись за составление алгоритмов машинного перевода с различных языков. У большинства не было машин, и правила перевода составлялись без учета возможностей техники и имели чисто теоретическое значение. Часть ученых «прогоняла» свои программы через машины. О. С. Кулагина, Е. В. Падучева, Т. М. Николаева, Н. Д. Андреев, Р. Г. Пиотровский и другие добились определенных успехов в разработке некоторых вопросов машинного перевода.



Одному из своих отчетов о работе над машинным, переводом американские ученые предпослали в качестве эпиграфа несколько иероглифов, передававших древнекитайское изречение, смысл которого заключался в том, что «для создания совершенного инструмента требуется время». К этому можно добавить «и труд».

Слабовольные люди, даже если они обладают недюжинными способностями, склонны легко «зажигаться» многообещающими идеями, но они редко доводят дело до конца. Кропотливая работа, растягивающаяся на многие годы, приводит их в уныние. Они берутся сразу за много дел, откладывая главное и убеждая себя, что им непременно надо сделать то-то и то-то, а уж после они засядут и завершат начатое. Проходит мно-

го времени, и они уже с недоумением глядят на листы, покрывающиеся легким налетом желтизны.

Машинный перевод оказался делом не для слабых. Известные ученые Г. Г. Белоногов и Р. Г. Пятровский так описывали то, что произошло с некоторыми бывшими «энтузиастами»:

«...После семи лет (1955—1961) исключительно интенсивных и оптимистических поисков — поисков, которыми занимались многие коллективы, наступил период значительного спада темпа и объема работ в области МП. Многие группы, прежде занимавшиеся вопросами МП, отошли от этой тематики, а часть таких объединений полностью распалась. Оптимизм и творческая фантазия сменились у некоторых пионеров МП унынием, и даже отрицанием разумности самой идеи МП».

Собственные неудачи часто выдавались за принципиальную невозможность создания экономически выгодного машинного перевода. Некоторые ученые утверждали, что качество машинного перевода никогда не достигнет качества переводов, выполненных квалифицированными переводчиками.

Но дело ведь было не в том, чтобы устраивать конкурентную борьбу между машиной и человеком. Машина должна была служить подспорьем человеку и, переводя (на первом этапе не так гладко, как человек) миллионы страниц технической и прочей информации, дать возможность специалистам хотя бы бегло ознакомиться с новыми сведениями и не упускать очень важные идеи. И если уж специалист заинтересуется важным для него сообщением, то он непременно постарается изучить все материалы, до последней запятой.

Машина не имеет еще «жизненного опыта» и знаний, на которые опирается в своей работе переводчик. Но вложенное в нее умение разбираться в значениях слов уже содержит в себе солидную толику опыта людей (переводчиков), создавших алгоритм.

СЛОВАРЬ

Я

зык медленно, но верно меняется. Некоторые слова устаревают, появляются новые. Развиваются наука и техника. Термины, употребляющиеся в новых отраслях, переключаются в повседневную речь. Только электричество подарило языку свыше 5 тысяч слов.

В английском языке сейчас насчитывается более 400 тысяч слов. В русском их около полумиллиона.

Конечно, в повседневной речи мы не употребляем и сотой их части. У образованного человека в обиходе от 3 до 5 тысяч слов. Даже великие писатели употребляли, сравнительно небольшое количество слов (Мильтон — 8 тысяч, Шекспир — 15 тысяч).

Для современных электронных машин «запомнить» даже сотни тысяч слов не составляет труда.

И даже сложнейшие логические построения оказались под силу машине, которая на любой предложенный ей вопрос отвечает только «да» или «нет». Так, например, весь процесс перевода надо было разложить на простейшие логические операции.

Сначала кажется, что вообще никаких элементарных логических операций нет. Есть, например, английское слово. Мы просто знали его эквивалент в русском языке, который и пишем, сообразуясь с грамматическими правилами и собственным здравым смыслом.

Но что такое здравый смысл? Не норма ли это для всякого живущего в обществе человека, которой он придерживается, чтобы быть понятым? А где есть норма, там должны быть логические правила.

Есть старая притча о сороконожке, которая всю жизнь бегала и никогда не задумывалась, какой из своих сорока ножек она ступает первой. Стоило ей задуматься над этим, и она навеки застыла на месте. Она не знала, с которой начать.

Итак, надо было узнать, с какой же ножки начинает свой бег многоногое насекомое. С какой логической операции начинает переводчик работу над текстом? Сотрудники лаборатории машинного перевода обратились к книгам и статьям теоретиков перевода и... ничего не нашли.

Я вспоминаю, с какой недоверчивостью отнеслись к идее машинного перевода не только переводчики, но и языковеды-грамматики. Они называли людей, занявшихся изучением логической структуры языка, двусмысленным словом «машинники», которое в их устах звучало почти как «мошенники».

Перед глазами переводчиков стояли все трудности, которые им приходилось преодолевать в процессе перевода. Иногда над переводом какой-нибудь одной фразы приходится думать часами. И кажется, что счастливое решение, ловкий и красивый перевод приходят неведомым путем. Но в эти часы проделывается огромная логическая работа.

Вновь и вновь переводчик вглядывается в сочетание слов, перебирая в уме все возможные значения их, он призывает на помощь все свои познания, свой личный жизненный опыт и многочисленные словари и справочники. Кажется, что решение может быть только одно — найти то, что написал бы по-русски автор переводимого текста. Но у разных людей получаются весьма отличные по своим оттенкам переводы одной и той же фразы. Осо-

бенно это заметно при переводе поэтических произведений.

Трудности завораживают переводчика, и порой ему кажется, что весь перевод состоит из подобных трудностей. Он забывает, что трудных мест бывает в среднем одно на несколько страниц, и не задумывается, как же он переводил то, что давалось ему легко и получалось гладко.

Когда переводчик становится теоретиком и садится писать статью о своем опыте, он чаще всего говорит об интересных и трудных случаях своей практики. Конечно, подразумевается, что читатели статьи получили начатки переводческой грамоты, изучая свой родной и иностранный языки.



Сотрудники лаборатории машинного перевода обратились к грамматикам, но нашли в них констатацию многочисленных фактов — формы слов и предложений и очень мало сведений о логической структуре языка. Как же оформляется мысль различными сочетаниями слов? Задача состояла в том, чтобы составить такую цепь логических операций, руководствуясь которой мог бы переводить человек, совершенно не знающий иностранного языка.

Очевидно, прежде всего машине понадобился бы словарь—перечень английских и русских слов. Это нужно было для того, чтобы машина могла узнавать английские слова и подбирать соответствующие им русские эквиваленты.

Самым простым решением было бы взять все имеющиеся английские толковые и англо-русские словари и ввести в память машины почти миллион английских и русских слов. Когда-нибудь электронные машины, которые будут производить миллиарды действий в секунду, вберут в себя весь словарный запас человечества. Но

пока такое решение было абсурдным. Машине потребовалось бы слишком много, как говорят, «машинного времени», чтобы отыскать нужное слово в подобном словаре. Для поиска лишь одного слова ей пришлось бы совершить невероятное количество логических операций.

Но этого делать и не требовалось. Дело в том, что примерно две тысячи слов — это тот словарный запас, которым мы обходимся, когда говорим. Две тысячи слов «покрывают» наши потребности в словах почти на 85 процентов. В остальные пятнадцать процентов может входить и сотня тысяч слов, но каждое из них употребляется нами редко.

Итак, одни слова мы употребляем очень часто, другие — редко. Подсчитано, что уже сто наиболее часто встречающихся слов входят в 20 процентов высказанных или написанных нами фраз. Словарь Пушкина, насчитывающий 21 197 слов, считается богатым.

Значит, надо было узнать, как часто встречаются те или иные слова, и взять для машинного словаря наиболее употребительные. Но как это сделать?

Нельзя сказать, что филологи прежде не занимались проблемой частоты встречаемости слов. В 1898 году появился первый частотный немецкий словарь Кединга. В 1911 году в США лингвист Элдридж составил словарь для иммигрантов из шести тысяч наиболее употребительных слов. Есть частотный словарь английского языка Торндайка.

В 1953 году в США на материале русских литературных источников XIX и начала XX века Йосселсоном был составлен частотный словарь русского языка. Но в нем большое место занимали архаизмы — устаревшие слова.

К 1963 году был составлен печатный частотный словарь русского языка на 2500 слов. Эта работа была проделана большой группой энтузиастов в Таллинском научно-исследовательском институте педагогики Эстонской

ССР. Сотрудники института несколько лет выписывали слова из небольшого текста в 400 тысяч слов на отдельные карточки, сортировали их. 2500 слов «покрывали» 80 процентов текста.

Защищена диссертация по частотности слов в произведениях Пушкина. Вручную было подсчитано, что в прозе поэт употребил слово «сказать» 659 раз, «один» 430 раз, «мочь» 393 раза, а в стихах — «День» 566 раз, «один» 529 раз, «любовь» 510 раз.

Частотные словари имеют большое практическое значение.

Для методики преподавания просто необходимо знать, какие слова и выражения в языке употребляются чаще других и какую роль они играют в речи. Знание частотности употребления слов дает возможность отобрать из них самые нужные. Выучив одну тысячу наиболее употребительных слов иностранного языка, учащийся может уже браться за чтение книг на иностранном языке. Он будет понимать большую часть текста, догадываясь и справляясь в словаре об остальном.

Частотность употребления слов необходимо знать связистам для борьбы с помехами, для составления кодов...

И создавая более или менее обстоятельный словарь для машинного перевода, необходимо было в первую очередь составить частотный словарь. Это была одна из первых серьезных проблем, с которыми столкнулась лаборатория машинного перевода. Важность этой проблемы трудно переоценить. Назревало решение, которое должно было дать возможность изучить и логическую структуру языка.

С унынием взирали сотрудники лаборатории на все растущую кипу текстов, которые следовало обработать, чтобы получить список наиболее часто употребляющихся слов и выражений. Конечно, можно было бы взять толстую книгу, как это потом сделали таллинские филоло-

ги, и вручную подсчитать, сколько раз встречаются все имеющиеся в ней слова. На это ушло бы несколько лет, и, пожалуй, такой словарь был бы годен для перевода только этой книги. Чтобы получить устойчивые статистические данные, требовалось обработать в десять раз больше текстов. Четыре миллиона слов! Египетский труд даже для большого коллектива исследователей.

Нужно было применить новые методы обработки текстов. И тогда руководители лаборатории приняли решение — будем работать по-современному... Тем более что новые методы обработки текстов позволили бы справиться и с другими проблемами. Например, проблемой многозначности...



Если вы заглянете в англо-русский словарь, то увидите, что одно какое-нибудь английское слово можно перевести самыми различными русскими словами. Некоторые словарные статьи даже не умещаются на одной странице, причем словарь часто дает лишь общее значение слова. Переводчику предоставляется возможность творчески подбирать синонимы в русском языке, а иной раз даже идти обходным путем и употреблять слова и целые выражения, которые на первый взгляд не имеют ничего общего со словами, стоящими в словарной статье. И тем не менее, ступив на этот обходный путь, переводчик довольно точно переводит английский текст.

В тексте многозначных слов обычно меньше, чем в словаре, где они составляют в среднем около 30% всех слов. В английских газетах из каждой сотни существительных — 21 многозначное. Для других частей речи числа таковы: прилагательных — 45, глаголов — 51, наречий — 21, предлогов и союзов — по одному многозначному.

Машина не даст сразу перевода многозначного слова. Ей придется еще много раз давать односложные от-

веты «да» и «нет», прежде чем в русском словаре найдется его точный переводческий эквивалент.

Машина ни о чем не может «догадаться», если мы не догадаемся сами, как надо переводить, если мы до тошно не проследим «ход» своих мыслей и не составим программу отдельных логических операций, которые приведут машину к тем же выводам.

Перевод слова, его эквивалент в русском языке зависят от того, в окружении каких слов стоит наше слово, иначе — от контекста. Так, русское слово «выдержать» может сочетаться го гловями «няпор ветра», «характер», «экзамен» и т. д., и всякий раз слово «выдержать» будет переводиться на английский другим словом.

В свое время, работая над словарем для переводческой машины, мы убеждались все больше и больше, что слово «само по себе» ничего не значит. Оно обретает жизнь, значение только в контексте, во фразе.

«Позвольте,— возразите вы,— а технические и научные термины?»

Да, когда-то термины были задуманы как слова, имеющие всего одно значение. Но даже они сегодня переосмысляются в различных науках, и если нам придется вводить научный текст, богатый терминами, в переводческую машину, то мы указываем, к какой отрасли знаний он относится — к физике ли, химии или строительному делу.

«Позвольте,— скажете вы,— иногда мы произносим всего одно слово, и оно понятно нашему собеседнику». Я вернусь к уже приведенному примеру. «Выдержит?» — спрашивает ваш собеседник. «Выдержит»,— отвечаете вы. Контекст слова заключается в ваших предыдущих словах, и вам обоим хорошо известно, будет ли стойким забор или ваш товарищ, сдающий экзамен.

К слову сказать, когда ребенок еще только учится говорить, он нередко связывает какое-нибудь явление с целой фразой, услышанной от взрослых. Он может не

знать слов, которые в нее входят, но уже употребляет эту фразу. Много позже он узнает и отдельные слова, встречая их в других фразах. А о существовании всяких грамматических категорий он узнает только в школе, уже свободно владея родным языком. Да и мы, взрослые, при беглом чтении иногда пытаемся схватить целиком смысл всей фразы, не разбивая ее на слова.

Иногда значение слова в какой-нибудь статье зависит от факта, приведенного в самом начале ее и отделенного от слова несколькими предложениями.

Все это мы подсознательно запоминаем и используем при переводе. Если мы невнимательны, то в нашем переводе неминуемы ошибки. Машина, которая имеет дело с огромным количеством ответственных переводов, не имеет права ошибаться.

И вот обыкновенная словарная статья иной раз разбухает до колоссальных размеров. Так, для того чтобы машина могла правильно переводить самое распространенное английское слово «be» — «быть», «являться», «иметь место» и т. д., и т. п., — при составлении логической схемы перевода этого многозначного слова пришлось предусмотреть около полутора тысяч логических операций. В результате слово «be» могло быть переведено в зависимости от английских слов, с которыми оно сочеталось, несколькими сотнями русских слов и выражений.

Для составления такой схемы пришлось рассмотреть 10 тысяч случаев употребления глагола «be». Фантастично, скажете вы. Действительно, сколько книг пришлось бы прочесть, чтобы выписать 10 тысяч примеров только на один глагол! А сколько времени потребовалось бы, чтобы классифицировать эти примеры, отобрать одинаковые и наиболее часто повторяющиеся!

Два миллиона слов, около 10 миллионов печатных знаков пришлось бы проанализировать исследователю

для составления подробных логических схем. Это многотомное собрание сочинений Чехова или Достоевского. Надо было изучить каждое слово и его окружение в этой громадине, чтобы появилась уверенность в том, что машина будет переводить правильно. На эту проблему понадобились бы десятки лет работы многочисленного коллектива лингвистов...

Было решено поручить работу по отбору словаря для машинного перевода искусным... счетно-перфорационным машинам. В то время подобная работа выполнялась ими впервые.

Для начала были отобраны самые различные тексты объемом в 2 миллиона слов. Девушки перфораторшницы набили их на специальные карточки. Для сортировочных машин была составлена программа действий, и работа началась.

Хлопотливо забегали держатели машин; тонкие щупы, попадая в отверстия карточек, давали сигналы управлять их в ту или иную пачку.

Рассортированные карточки попадали в быстродействующее печатающее устройство, из которого довольно быстро выползала длинная широкая бумажная лента.

Получив сотни рулонов бумажной ленты, мы могли убедиться, что все наши пожелания были выполнены превосходно.

Так что же входило в программу действий счетно-перфорационных машин? Какую работу с английскими текстами должны были проделать эти механические логики?

Нам надо было знать:

какие слова употребляются в английских текстах и сколько их;

какие слова употребляются чаще других и сколько раз;

в какие английские выражения входят эти слова и вообще с какими другими словами они сочетаются;

сколько слов состоит из одной буквы, сколько из двух, трех и так далее?

И мы приказали машинам:

рассортировать слова по числу букв, чтобы сначала шли слова однобуквенные, потом двухбуквенные, трехбуквенные и т. д.;

внутри каждой такой группы выстроить слова по алфавиту от «а» до «зет»;

все одинаковые по написанию слова сгруппировать и выстроить в 'столбики;

справа и слева от каждого слова дать по небольшому куску текста, чтобы можно было анализировать слова в окружении, в контексте, то есть составить словарь — конкорданс.

И это еще не все. Чаще всего значение одинаковых по написанию слов зависит от слов, следующих после них (вспомним, «выдержать характер», но «выдержать экзамен»). И вот каждое первое слово куска текста, прилегающего к основному слову справа, тоже было рассортировано по алфавиту, и все одинаковые выражения, идиомы, устойчивые сочетания, в которые входило основное слово, оказались сгруппированными в определенных местах.

А для выявления характерных форм каждого слова машины провели сортировку слов по окончаниям. Грамматические явления как бы подавались на тарелочке. Это была мечта лингвиста!

Сразу стало наглядным — какие слова и выражения употребляются чаще (машины даже сами подсчитали многие интересные статистические данные и отпечатали их в конце рулонов), какие сочетания слов в языке возможны, а какие нет.

— Ого! — воскликнул кто-то, разворачивая рулоны.— Да тут добрая сотня кандидатских диссертаций в готовом виде.

И, пожалуй, он не ошибся. Годами собирают аспиран-

ты — языковеды и литературоведы — подобный материал. Вручную подсчитывают число глаголов и существительных в произведениях Толстого, Голсуорси и Шолохова. Машина могла бы сделать подобную работу в очень короткое время.



К лету 1957 года было решено, что для машинного словаря надо взять 16 тысяч слов. Впоследствии, когда через счетно-перфорационные машины «прогнали» еще миллионы слов, был окончательно составлен словарь, который насчитывал 13,5 тысячи английских слов и около 40 тысяч русских слов и выражений.

Было решено включать в словарь все слова, встречающиеся среди почти пяти миллионов слов текста чаще двух раз, и многие слова, попадавшие два или один раз.

Такой словарь «покрывает» 98,5 процента английско-го текста. Практически это означает, что машине дана возможность «понимать» и переводить 99 слов из ста при чтении любой английской и американской газеты. Один процент падает на имена собственные, которые было решено оставить в их латинском написании, и лишь полпроцента или одно из каждых двухсот слов может оказаться машине неизвестным.

Это очень высокая степень надежности. Анализ дополнительных текстов еще на миллионы слов дал бы очень мало, ибо в отборе слов выявилась интересная закономерность.

Предел — «покрытие» словарем всех ста процентов текста недостижим, как скорость света. Можно без конца приближаться к пределу, но язык в каждый данный момент может родить новое слово. Каждый новый десяток тысяч слов, включенный в словарь, увеличивал бы надежность перевода лишь на ничтожную долю процента.

Одна тысяча английских слов занимает 70 процентов текста. Две тысячи — 87 процентов. Десять тысяч — 97 процентов. А двадцать две тысячи — 98,5 процента. Вот как невелик выигрыш при увеличении словаря более чем в два раза.

Для научных текстов (например, для одной из областей математики) надежность достигается и при меньшем словаре. Так, оказалось, что три тысячи слов (из которых 700 многозначных) покрывают 97 процентов текста.

Как я уже говорил, проще было бы строить словарь на основе существующих переводных и толковых словарей. Но оказалось, что обычные переводные словари не дают частоты встречаемости каждого слова. Эти словари включают множество слов самых различных жанров, стилей, эпох, пытаясь охватить все, что накопил язык за многовековое развитие. Многие слова ныне употребляются крайне редко. И в то же время в словарях нет еще многих слов и выражений живого языка, уже встречающихся в литературе, в газетах и журналах...

Первая тысяча слов нашего словаря встречалась наиболее часто. Восемьдесят слов из каждой сотни слов текста входили в эту тысячу. И более половины из них составляли знаки препинания (которые было принято считать словами), предлоги, вспомогательные глаголы, артикли, глаголы «be», «do», «make» во всех формах. Собранные в группы, они стали нашей опорой уже в самом начале работы, своеобразной сетью геодезических пунктов, к которым топографы «привязывают» местность во время работы.

Проведя большую подготовительную работу по анализу английских текстов и составлению словаря, можно было приниматься за создание алгоритма машинного перевода с английского языка на русский.

Ч

Чтывая газетные статьи о том, как машины считают, планируют, переводят, нередко поражаешься той легкости, с какой авторы их расправляются с труднейшими задачами. Стоит появиться во фразе слову «машина», как сейчас же все трудности оказываются разрешенными, а ученым остается только сидеть сложа руки или улыбаться набежавшим журналистам.

Быть может, в этом виноваты сами ученые, очень неохотно рассказывающие о бессонных ночах, когда чистый лист бумаги покрывается завитушками, а идея ускользает, хотя всего час назад, пока ехал в троллейбусе, все было ясно и аргументировано. Какой-нибудь упрямый факт, пришедший в голову в последний момент, разрушает стройную систему умозаключений.

Да и трудно рассказывать о творческой кухне, о том, как идея обрела плоть, как проходили бесплодные дни самообвинений в бездарности и в то же время совершался «таинственный» процесс кристаллизации фактов вплоть до блаженной минуты «озарения».

И это еще только начало. Нужно убедить всех в целесообразности осуществления своей идеи, нужно организовать работу.

Иногда работа предусмотрена планами, и ее просто поручают сделать. Но от этого несколько не легче. Кто и когда измерил груз ответственности человека, в кото-

рого верят и на которого надеются? И потом оказывается, что объем работы растет на глазах. Он растет подобно цепной реакции. Бывают минуты страха, как перед джинном, выпущенным из бутылки и вытекающим из ее горлышка длиннющей стружкой дыма, которая становится живой глыбой величиной с многоэтажный дом.

Но вот проходит горячка первой поры. Уже можно окинуть взглядом эту глыбу и подсчитать, сколько потребуется времени и сил, чтобы справиться с ней.

Жестоко ошибается тот, кто думает, что труд ученого состоит из бесконечных взлетов, падений и неожиданных провалов, за которыми следуют новые взлеты. Наступают рабочие будни, когда целые годы уходят на подбор фактов, часто неинтересные, но необходимые опыты, бесчисленные проверки, на механическую работу, которая требует не слишком большой квалификации. От этого никуда не денешься. Монотонность, рутинность — и все время мысль: «А что же в конце концов получится?»

Одни и те же операции повторяются вновь и вновь. Хорошо бы приспособить машину для их выполнения! Но на составление такой программы тоже уйдут годы. Надо сначала закончить то, что задумано.

И, наконец, все уложено в стройные схемы, и наступает акт драмы, именуемый составлением алгоритма, то есть совокупности правил, благодаря которым к машине должна перейти хотя бы небольшая часть способностей, присущих до этого только человеку.



Шли годы. Постепенно вырисовывались все детали алгоритма машинного перевода. Словарь машины. Четыре тысячи логических схем многозначных слов. Двести грамматических схем. Схемы отбрасывания грамматических окончаний, схемы различения омографов, грамматические таблицы русских слов, схемы синтеза русской фразы... И в каждой схеме десятки и даже сот-

ни вопросов, на которые машина по программе должна отвечать «да» или «нет».

Сколько же знает даже средний переводчик! Как необыкновенно сложен механизм лишь одной из сторон человеческой деятельности! Чтобы составить программу ее, десятки людей трудятся годы и годы, и при этом их не покидает ощущение, что они как-то упрощают процесс перевода, что они не могут распознать его до конца. Но многое уже получается.

Человек не переводит слово за словом. Он думает о законах того языка, на который переводит, он думает о стилистике, о красоте слога.

Вот простая английская фраза: «Last night I went to bed at ten o'clock». Человек неискушенный может перевести ее дословно: «Последней ночью я пошел в постель в десять часов». Перечитав русское предложение, он поймет — что-то здесь не так. И, немного подумав, напишет, наверное, его заново: «Вчера вечером я лег спать в десять часов».

Программа для машины должна предусматривать грамотный и красивый перевод.

Подготовлены все схемы, и по ним «прогоняют» английские фразы. Прежде чем алгоритм будет введен в машину, снова и снова устраиваются проверки — «тесты». Этим могут заниматься люди, совершенно не знающие английского языка, — схемы должны автоматически выдать правильный перевод на русский.

Приходит время отладки программы и ее первых испытаний на машине...

Никакая схема и даже перечисление трудностей не даст представления об огромном напряжении духовных и физических сил, которое требуется от людей на завершающем этапе работы. Надо было самому побывать в то время в лаборатории и посмотреть на усталые лица людей, лихорадочно выискивающих ошибки и вносящих последние поправки.

Работа над машинным переводом разбита на две большие стадии. Первую из них можно было бы грубо охарактеризовать как лингвистическую, вторую — как собственно «машинную».

Во время первой стадии лингвисты — преимущественно переводчики и специалисты в области русского языка — стараются изложить на бумаге лексические и грамматические соображения, которые возникают у них в процессе понимания английской и построения русской фразы. Делают они это в виде пространных схем, таблиц и списков слов, ограничиваясь в своей работе основным принципом классической логики. Эта логика двузначна и знает лишь две оценки суждений: «истинно» и «ложно».

В физике, при рассмотрении некоторых явлений, встречается и третья оценка — «неопределенно». Но нам, желающим во что бы то ни стало получить ответ «да» или «нет», такая оценка не подходит. Мы без конца задаем вопросы: «Сочетается ли такое-то слово со словом таким-то?», «Имеет ли слово такую-то форму?» и т. д., и на основании полученного ответа пишем новый элемент правил перевода или продолжаем задавать все новые и новые вопросы.

Вторая стадия заключается в программировании полученных схем, в продумывании последовательности их работы и в отладке программ уже на самой машине. Здесь главную роль играют математики-программисты и математики-логики, знатоки электронных машин и те лингвисты, которым настолько понравилась математика и техника, что они в процессе работы усвоили программу специального высшего учебного заведения, как, например, Юрий Николаевич Марчук, ныне один из руководителей лаборатории, доцент Московского педагогического института иностранных языков и соавтор курса «Вычислительной лингвистики».

Если мы, хотя бы в качестве зрителей, приняли бы участие в испытаниях алгоритма машинного перевода, то, глядя на пульт быстродействующей электронной вычислительной машины, мы, конечно, увидели бы немного. Нам покажут английский текст, перфорированные ленты, стойки, на которых смонтированы элементы машины, буквопечатающее устройство, из которого выполняет лента с русским переводом...

Что же происходит там, за дверцами металлических шкафов, в недрах «электронного мозга»? Об этом можно рассказать... Ради этого трудились долгие годы десятки людей. Последовательность выполнения машиной логических операций не совпадает с последовательностью работы над алгоритмом.

Но рассказывая о работе машины, можно понятней и проще рассказать о работе людей и возвращаться к той или иной задаче, которая сначала ставила в тупик сотрудников лаборатории. Мучительные порой поиски увенчались успехом, и небывалый по сложности алгоритм должен был «сказать свое слово»...



Сейчас, наверно, не найдется человека, который не вырисовывал бы печатными буквами почтовые индексы на конвертах. Наша почтовая система вводит сортировочные машины, сложные оптико-электронные устройства, способные читать печатные цифры и действовать в соответствии с заключенной в них информацией.

Считывающее устройство — это глаза и машины-переводчика. Но пока она считывает не печатный текст, а сочетания дырочек на бумажной ленте, пробегающей перед фотоэлементом. Сочетания дырочек — это уже числа в двоичной системе. Английский алфавит передается цифровым порядковым кодом (1, 2, 3, 4, 5 и т. д.) Английский текст предварительно набивали на ленту

специальными машинами — перфораторами, которые букву «а», например, заменяли единицей (01 в двоичном коде), «b» — двойкой и т. д, до «z» — 26. Также вводилась в машину до этого и вся программа ее работы.

Казалось, что все идет хорошо. Но уже в том, что текст вводился в машину при помощи ленты, заключена проблема огромной важности.

Когда было принято решение создавать «промышленный» алгоритм для перевода с английского языка, на котором во всем мире публикуется большая часть иностранной научной и прочей информации, сотрудники лаборатории в первую очередь занялись экономикой машинного перевода.

Они делали расчеты, сравнивали возможности человека и машины, изучали экономические выкладки иностранных коллег. Американцы подсчитали, что в то время машинный перевод каждого слова обошелся бы в 1,46 цента, причем поиск слова в словаре, логическая обработка его, печатание стоили бы всего 0,6 цента, а вот подготовка текста, набивка его на ленты обошлась бы в 0,86 цента за слово. Американские расчеты подтвердились. Самым дорогим оказался ввод данных в машину, и это ставило под угрозу идею машинного перевода.

С годами эта цифра сократилась в несколько раз, но решение проблемы продолжало зависеть от систем ввода текста в машину.

Действительно, в вычислительных центрах можно увидеть десятки и даже сотни девушек, сидящих у перфорационных машин. Они нажимают клавиши устройств, похожих на пишущие машинки, и перфораторы пробивают отверстия в бумажной ленте или в карточках. Затем перфорированная лента помещается в вводное устройство и пробегает со скоростью более метра в секунду под «глазом» машины — фотоэлементом, посылающим в «память» импульсы тока.

Так вводились в машину все сведения, нужные ей

для работы-, включая алгоритм — совокупность программ, систему команд, определяющих последовательность действий.

Такой способ ввода был явно непригоден для машин, с каждым годом работавших все быстрее и быстрее. В одном из иностранных журналов даже подсчитали, что современная переводческая машина могла бы прочесть 1800 тысяч букв в минуту, но, чтобы снабдить ее перфорированными карточками, понадобилось бы 12 тысяч машинисток, работающих со скоростью 10 тысяч букв в час. Кроме того, для проверки и редактирования потребовалось бы 10—12 тысяч сверщиков и еще столько же машинисток. По мнению журнала, для обслуживания пднпй машины для переводов понадобился бы целый город с населением в 50—100 тысяч человек.

Мнение неутешительное, но... ученые многих стран уже работали над устройствами, благодаря которым машина могла бы свободно читать печатный текст.

Было проведено немало удачных и неудачных опытов, прежде чем ученые разработали автоматы, которые сами распознают буквы, цифры и другие символы и преобразуют их в двоичный код.

Сначала такие автоматы стали применяться для чтения цифр на банковых чеках. Потом был создан автомат «Эра», который со скоростью 120 знаков в секунду читал текст, напечатанный на пишущей машинке.

Ученые исходили из того, что в нашем мозгу записаны эталоны-контуры знаков и цифр, что позволяет сравнивать и узнавать увиденное.

Для чтения была использована катодно-лучевая трубка, которая применяется в телевидении. Обегающий букву луч света отражается с различной интенсивностью в зависимости от того, попадает ли он на чистое место бумаги или на часть знака. Отраженный свет фокусируется и подается в специальный «прибор опознания»,

в котором каждый распознанный знак превращается в импульсы, передающие его в машину закодированным по двоичной системе. Однако и этот способ оказался далеко не совершенным.

Существует большое количество типографских шрифтов, а различных почерков — не счесть. «Эра» могла читать текст, напечатанный только определенным шрифтом. Но человек узнает букву, как бы она ни была написана.

В настоящее время разработаны более совершенные системы считывания текста. В частности, использовался статистический подход к опознаванию знаков.

Ученые обратили внимание на то, как человек распознает предметы. Очевидно в мозгу все-таки хранятся не эталоны предметов, а их обобщенные образы. Всех клеток мозга не хватило бы, если бы ему пришлось создавать по эталону на каждый чуть отличающийся от другого предмет. Мать показывает сынишке елку. Образ ее запечатлевается у него в памяти без объяснений. И потом, уже встретив елку побольше или поменьше, густую или тощую, он безошибочно называет ее.

Буква «б», например, имеет нечто общее в самых различных шрифтах. Это «нечто» — кружок и закорючка, начинающаяся в верхней левой части кружка и идущая вверх и направо. Но буквы «б» из разных шрифтов могут совпасть далеко не во всех точках, если мы их попытаемся совместить.

Ученые определили «статистическую меру сходств» букв из разных шрифтов и как бы создали обобщенный образ каждой буквы алфавита. Они установили, что точки совмещенных букв лягут гуще в определенных областях пространства. Эти кучно лежащие точки и будут создавать образ, который поможет распознавать все новые и новые варианты одних и тех же букв.

Есть автоматы для чтения, работающие и на других принципах. Во всяком случае, «видит» машина уже неплохо.

Итак, мы вводим в машину английский текст, каждая буква которого специальной программой обозначается числом — от 1 до 26. В каждом введенном слове определяется число букв, так как в словаре машины все слова тоже сгруппированы по числу букв (например, все четырехбуквенные слова сгруппированы по алфавиту в одном месте). И сразу же машина начинает искать слово в словаре, где все буквы закодированы теми же числами. Слово «age» выглядит в машине так: 1, 7, 5.

Но одновременно с поиском слова в словаре работает программа исправления искажений. Дело в том, что считывающий автомат пока еще читает правильно только 98 процентов текста, а это значит, что определенное количество букв в словах может быть искажено. Да и при вводе текста с перфорированной ленты могут быть ошибки, так как от них не застрахована даже опытная перфораторщица.

Короче говоря, машина принимается выполнять обязанности корректора, который в любом издательстве вылавливает и исправляет ошибки в корректурных оттисках, перед тем как ротационные машины начнут безостановочно печатать весь тираж.

Как же машина справляется с такой работой? В программе ее предусмотрено много способов исправления искажений. Она может обратиться к своему словарю, найти там слово, похожее больше других на наше искаженное слово.

Возьмем для примера русское слово «котолый». Вы скажете, что такого русского слова нет. Я напомню вам, что слово искажено, и вы тогда сразу догадаетесь, что в слове ошибка, и станете читать «который». Вы определили это простым сопоставлением букв искаженного и неискаженного слова. Пример нарочно взят легкий, но он позволит серьезно поговорить о некоторых по-

ложениях теории информации, применимых на практике.

Один из основоположников теории информации К. Шеннон считает речь разновидностью вероятностного процесса. Наша речь — это последовательность определенного числа элементов. И в каждом определенном месте этой последовательности может оказаться любой из элементов. (Элементами Шеннон считает буквы.) Но вот вероятность появления того или иного элемента в определенном месте для разных элементов разная. Для иных она даже нулевая (например, в русском языке между буквами «ый» не может появиться согласная). И вообще вероятность появления каждого следующего элемента во многом зависит от сочетания предыдущих.

Для создания программы исправления искажений необходимо было изучить закономерности появления в том или ином месте тех или иных элементов (букв).

Письменный язык состоит из небольшого числа элементов (английский из 26), и появление их предсказуемо. Мы знаем частоту встречаемости букв в английском языке на каждую тысячу букв текста:

Буква	Частота встречаемости	Буква	Частота встречаемости	Буква	Частота встречаемости	Буква	Частота встречаемости
E	131	I	63	C	28	B	14
T	105	S	61	M	25	V	9,2
A	86	H	53	U	25	K	4,2
O	80	D	38	G	20	X	1,7
N	71	L	34	Y	20	J	1,3
R	68	F	29	P	20	Q	1,2
				W	15	Z	0,77

Теория информации возникла из практических нужд— ученые разработали ее, чтобы выяснить, какой код лучше всего подходит для сообщений по телеграфу, как

избавиться от искажений при передаче сигналов. Но затем теория информации перекочевала и в другие науки — кибернетику, лингвистику, биологию, и даже есть попытки применить ее в эстетике.

Информация, как и мысль, не бесплотна. Для того чтобы передать сообщение, нужна какая-то система знаков, звуков, сигналов, понятная всем тем, для кого предназначено сообщение. Таких систем на свете множество. Их называют кодами. Многие ученые считают кодами языки и даже изобразительные и другие средства, которыми пользуется искусство.

При равновероятностном, совершенно случайном появлении любой из букв алфавита текст выглядел бы совершенно случайным набором букв:

БКСХЪЕ КФГОРБУХЗТЧ ЫАПРЖДЛЙЦУХЗЩ

Но раз язык — система, то такого положения в нем быть не может. Человек не свободен в выборе каждой следующей буквы, когда он пишет. Так, в английском языке за буквой «Q» всегда следует «U».

Следовательно, для исправления искажений нужно было составить таблицы, показывающие наиболее вероятные комбинации двух и трех букв. Если машина не находила в таблицах встретившееся ей сочетание, то она меняла его на наиболее вероятное, и статистика показывала, что почти на сто процентов текст оказывался исправленным.

Но устойчивые комбинации букв встречаются в языке не очень часто, и тогда в исправлении ошибок очень помогает избыточность языка. Она заключается в том, что в языке употребляется больше элементов, чем это нужно для какого-нибудь сообщения.

«Изменив цифру,— пишет известный американский ученый Дж. Пирс,— мы получим новое число, но если мы изменим букву или даже несколько букв в тексте обычного сообщения, то скорее получим искаженное, но

могущее быть узнанным, чем другое осмысленное сообщение. Буквальный же смысл искаженного текста будет бессмыслицей. Это является весьма ценным свойством письменной речи. Благодаря этому свойству мы оказываемся в состоянии, например, понимать плохой почерк. Та же избыточность позволяет нам уловить смысл плохо расслышанной фразы и понимать человека, который в разговоре грамматически неправильно употребляет слова или говорит с иностранным акцентом».

Из-за избыточности плотность информации на каждую букву языка очень мала. Но именно это помогает нам избегать многих ошибок и увеличивает не только число признаков появления той или иной буквы, но и число признаков, определяющих значение каждого слова.

Слова в языке длиннее, чем они могли бы быть. У нас есть слова из десяти букв. А великое множество трехбуквенных сочетаний нами не используется. Чем не хороши слова — «коа», «эню», «лок»? И таких трехбуквенных сочетаний теоретически можно составить тридцать тысяч. А в русском языке употребляется всего несколько сот трехбуквенных слов. Каждый язык использует свои, присущие ему не только трехбуквенные, но и четырех-, пяти- и т. д. буквенные сочетания.

Например, в русском языке довольно редко приходится на одно слово сразу четыре согласных, не разделенных гласными (слова типа «монстр всплыл»).

Далеко не каждая буква слова несет информационную нагрузку. Представьте себе, что вы нашли клочок газеты, на котором есть такие буквы: «крепить обороно-», а дальше оторвано. Вы легко догадаетесь, что там должно быть «крепить обороноспособность». В чем тут дело? В вашей гениальности? Нет. Просто язык обладает избыточностью. Избыточность помогает нам понимать искаженные телеграммы.

Эта избыточность создается грамматическими правилами, которые услужливо подставляют к основам слов

те окончания, которые мы ожидаем увидеть. Эта избыточность создается определенным порядком слов, узаконенным многократным употреблением некоторых фраз. Нам будут понятны даже такие обрывки фраз в искаженной телеграмме: «Поздр....нем рожд... же....ровья и ...частья. Иванов». («Поздравляю с днем рождения, желаю здоровья и счастья. Иванов»), хотя никто и никогда так не писал.

Перед нами осмысленный текст, обусловленный явлениями реальной действительности. На этом основано и определение количества информации по методу, предложенному Шенноном. Он показывал английский текст частично, до некоторой буквы, и предлагал угадать следующую букву. После этого показывалась действительная следующая буква текста и предлагалось угадать следующую за ней букву и т. д. Текст содержал 129 букв, из них 89 были угаданы верно, что составляет 69 процентов от всего текста.

Избыточность языка позволила Шампольону расшифровать письма древних египтян. Вспомним, как расшифровывали размытую записку в «Детях капитана Гранта» или записку о сокровищах в «Золотом жуке».

Для удобства работы алгоритма считается, что всякое новое слово, введенное в машину и отсутствующее в словаре, искажено. При несовпадении одного знака с имеющимися в словаре словами орфографическая ошибка исправляется. Чем длиннее слово, тем больше букв в нем совпадет со словом в словаре и тем легче исправить искажение.

Избыточность позволяет «свертывать» или «сжимать» слова для экономии места в памяти машины. Пример свертывания: представление имени и отчества инициалами. В старину на Руси определенные слова писали «под титлами», опуская некоторые сочетания букв и обозначая их условными надстрочными знаками.

Установлено, что все естественные языки имеют большую избыточность (с точки зрения теории информации). Ее можно измерить. Для основных европейских языков она составляет около 50—80 процентов. В практике машинного перевода можно заменять слова аббревиатурами, отбрасывать некоторые буквы и т. д.



Как я уже говорил, одновременно идет поиск в словаре всех введенных в машину слов. Но машинный словарь имеет свои особенности. Они заключаются в том, что все слова в нем без окончаний. Английские существительные стоят в единственном числе, глаголы—в первом лице, а прилагательные не имеют степеней сравнения. И только неправильные глаголы имеют все формы.

А мы ввели в машину текст, и слова в нем имеют окончания. И тотчас начинает работать программа отбрасывания окончаний. Слова как бы обрезаются, а окончания отбрасываются, но не совсем. Их направляют в определенное место в памяти машины, чтобы «сбегать» за ними, когда в них будет нужда.

У каждого слова в машине есть своя «ячейка», которая состоит из нескольких десятков «разрядов» (обычно это крохотные элементики, способные удерживать положительный и отрицательный заряды, «ноль» или «единицу», из которых складываются числа). В «ячейку» записывается слово, его номер, грамматическая и прочая информация о слове, и вообще все сведения, полученные в результате работы каждого этапа алгоритма.

Слова без окончаний снова проверяются по словарю, получают здесь свой (не перевод, нет, до перевода еще далеко) цифровой эквивалент. Само слово, закодированное числами, как бы исчезает, и теперь машина имеет дело до самого конца с его цифровым эквивалентом.

Цифровые эквиваленты слов сыграли большую роль в накоплении информации о каждом слове. Решение это

пришло уже после того, как были составлены многие схемы. Схемы были громоздки, они содержали все сведения о слове. Работники лаборатории машинного перевода подумали: а что, если сам номер слова будет, говорить о том, какая это часть речи, многозначно слово или однозначно?.. Предположим, что мы взяли 50 тысяч номеров и с десятого по десяти тысячный отвели для обозначения однозначных существительных. А если номера с десятого по шеститысячный мы отведем для имен существительных неодушевленных, а номера с 1000 до 2000 для географических названий, причем номера с 1500 по 1600 присвоим столицам государств? Значит, английское слово «Moscow» — «Москва» получит, к примеру, номер 1525, который сам по себе говорит о том, что это однозначное существительное, неодушевленное, географическое название, столица. Число сведений, которые оказались возможным передать с помощью номера, практически не ограничено.

Номером (или цифровым эквивалентом) можно, например, передать тот факт, что данное существительное означает должность, организацию, деньги, документ, группу людей, собравшихся в одном месте, и т. д. Или что оно — числительное (в машинном переводе нет числительных: по своим формальным признакам все количественные числительные стали существительными, а порядковые — прилагательными). Номер сообщает о глаголе, что он модальный, или вспомогательный, или глагол движения и т. д. С помощью номеров мы сразу же откладываем про запас сведения о любой части речи.

В общем цифровой эквивалент содержит все сведения, органично присущие данному слову, независимо от языка и контекста¹. На эти сведения мы будем опираться

¹ В тот период способ кодирования классов объектов путем присвоения элементам этих классов порядковых номеров из заданных числовых интервалов был уже известен в теории и практике применения счетно-перфорационных машин. В машинном переводе

ся в дальнейшей работе алгоритма и особенно, когда придет пора согласовывать все русские слова уже в русской фразе.

И снова «но». Оказалось, что многим словам, входящим в словарь, нельзя давать цифрового эквивалента. 3200 слов из 13,5 тысячи слов словаря оказались омографами. Они получают для начала служебные эквиваленты, так как неизвестно даже, что это за части речи. Специальный служебный эквивалент получают и слова новые, не учтенные в словаре.



В английском языке слова, совершенно одинаковые по написанию, очень часто выступают в роли одновременно и существительного, и глагола, и других частей речи. В русском языке это явление встречается довольно редко. По-английски «work» — «работа» и «work» — «работать». Для незнающих английского языка будет понятен пример: «печь» — существительное и «печь» — глагол. Богатство суффиксов и окончаний в русском языке позволяет избежать одинакового написания. Но даже если слова и одинаковы по написанию, мы безошибочно определяем, к каким частям речи они относятся.

Как мы это делаем? По контексту, по окружающим это слово другим словам, по десяткам признаков, которые мы находим в тексте. Для вас не составит труда определить, что «печь» в сочетании со словами «хлеб, картофель» непременно будет глаголом.

Точно так же работают и схемы различения омографов в алгоритме машинного перевода.

Взгляды лингвистов на определение омографов расходятся. Но мы не станем вдаваться в ученые споры

такая практика сыграла свою положительную роль. Но преимущества этого способа обращаются в недостатки, как только объем словаря превышает запланированный. В настоящее время идут поиски новой системы.

и будем исходить только из практических нужд машинного перевода. Для нас омографы — это разные слова, имеющие одинаковое написание.

Три тысячи двести омографов. Не определив, к какой части речи относится каждый из них, мы даже не можем обратиться к словарю. Действительно, как можно сказать, что значит слово «печь», не зная даже, существительное это или глагол? А в английском языке есть слово «round», которое может быть одновременно существительным, глаголом, прилагательным, наречием и предлогом. Это слово можно перевести русскими словами — круг, округлять, круглый, вокруг и за (углом).

Есть много способов определить часть речи такого слова. Если подводит один способ, мы обращаемся к другому. Слово «works» может быть и глаголом в третьем лице («работает») и существительным во множественном числе («работы»), о чем говорит такой признак, как окончание «s». Но в тексте у нас стоит «he works». Проверив, является ли слово, стоящее слева, подлежащим, и убедившись в этом, мы твердо говорим, что это глагол, и отправляемся в машинный словарь за получением цифрового эквивалента, а впоследствии и перевода.

Но омография не исчерпывается примерами, которые мы привели. В русском языке встречаются любопытные омографы: «три» (число) и «три» (от «тереть»), «стекла» (сущ.) и «стекла» (глагол.), «лечу» (от «лететь») и «лечу» (от «лечить»). В английском языке примеров омографии во много раз больше. Там могут совпадать как формы одного и того же глагола («put», «put», «put»), так и формы разных глаголов («found» (от «find») — «нашел» и «found» — «основать»).

Интересна схема разрешения синтаксической омографии, или схема сложных частей речи. В английском языке, как и в русском, есть тысячи устойчивых сочетаний слов. Вот, например, английское «by all means» или русское «во что бы то ни стало». Все вместе слова, входя-

щие в сочетание, выполняют обычно функцию одного члена предложения и даже одной части речи (здесь — наречия). Мы не переводим такое сочетание слово за словом, а даем перевод русским устойчивым сочетанием. И если бы мы взглянули в словарь, то увидели бы, что против каждого слова английского сочетания стоят совсем другие русские слова. Поэтому мы считаем, что эти слова только совпадают со словами словаря по написанию и являются омографами. Для нас все сочетание — как бы одно слово. Это очень удобно, потому что уже в машинном словаре мы можем предусмотреть для очень многих английских выражений красивый, добротный русский перевод.

Так в приводившемся уже мною примере «Last night I went to bed at ten o'clock» сочетание «last night» переведется не буквально «последней ночью», а «вчера вечером». В одном из учебников языкознания говорится, что «не следует смешивать омонимию с многозначностью слов». Так, по учебникам слова «лук» (растение) и «лук» (оружие) будут омографами. Но машинная лингвистика часто пренебрегает правилами, созданными традиционной лингвистикой. Для удобства оба эти слова считаются одним многозначным словом, многочисленные значения которого могут быть выяснены в результате работы специальной схемы. Но мы забегаем вперед, ибо время работы программы многозначных слов еще не пришло.

Как видите, мы, собственно, не приступали к работе алгоритма и еще «копаемся» в словаре, а уже пошли необычайные сложности.

Если бы мы захотели показать, как переводится в машине только одна фраза и какой логической обработке подвергается каждое слово, то для этого понадобился бы пухлый том с массивным приложением в виде графиков и схем.

Среди них была бы схема, которая определяла бы, какой частью речи является любое слово, если даже его нет в нашем автоматическом словаре. По ней машина будет действовать, как те студенты, которым языковед Лев Владимирович Шерба предложил на вступительной лекции по языкознанию разобрать по частям речи фразу: «Глокая куздра штеко будланула бокра и курдячит бокренка».

По чисто формальным признакам мы можем сказать, где здесь существительное, а где глагол. Мы можем определить род, число, падеж, наклонение и многое другое. В английском языке к формальным признакам относится место слова в предложении, его окончание, сочетаемость с предлогами и другими словами. Вот такую «куздру», даже не зная, что она значит, уже можно запускать в машину и опираться на нее в грамматических схемах.

Очень важным в машинном переводе оказалось знание возможностей и работы быстродействующих электронных вычислительных машин. Можно придумывать какие угодно правила перевода, но если их нельзя «реализовать» на машине, то грош им цена.

Когда сотрудники лаборатории машинного перевода создали лексические, грамматические и прочие схемы перевода, оказалось, что для составления программы алгоритма потребуется... 15 миллионов команд, то есть указаний машине, где какие данные взять и в какое место направить, произведя то или иное логическое действие. Это 250 миллионов битов информации, и для запоминания ее потребовалось бы 15 миллионов «ячеек».

Память для этого нужна колоссальная, не говоря уже о том, что программировать такой алгоритм очень тяжело. Надо было бы привлечь к работе сотни программистов. А у каждого программиста своя манера работать, да и число ошибок при программировании было бы велико, потому что людям свойственно ошибаться.

ся. И одна ошибка может привести к сбою программы и застопорить работу всего алгоритма.

Это стало ясно примерно в середине работы над схемами перевода. Мы, лингвисты, чертили десятки тысяч блоков, указывая в них, какие логические действия делать машине. И мы замечали, что часто повторяемся, задавая вопросы и требуя на них ответа «да» или «нет». С каждым годом работы беспокойство росло. Начались разговоры, что машина не «потянет» и наш громоздкий алгоритм ляжет тяжелым и бесполезным грузом на плечи государственного бюджета.

Встал вопрос, а нет ли в наших схемах одинаковых операций? Нельзя ли унифицировать команды и значительно уменьшить их число? Выход был найден, и 15 миллионов команд свели к пятнадцати тысячам.

Заслуга эта никак не могла принадлежать лингвистам. Поистине революционную работу проделали математики и логики. В свое время эту идею выдвинула математик О. С. Кулагина.

Так появились операторы, являющиеся существенной частью того, что теперь называют входным языком (или машинным языком).

Восемь операторов дают возможность производить любые проверки, сравнения, лингвистические операции.

С помощью одного оператора проверяют, нет ли в тексте какого-нибудь конкретного слова справа или слева от нашего слова (или списка слов — например, глаголов движения).

Этот оператор очень важен. Вот две русские фразы: «Я был в городе» и «Я пошел в город». Здесь, несмотря на один и тот же предлог «в», слово «город» стоит в разных падежах. Во втором случае написание «город» обуславливается тем, что перед предлогом стоит глагол движения.

Для того чтобы выяснить значение слова, важно знать его окружение. (Вспомните: «выдержать экзамен,

характер...») Перевод часто зависит от вхождения нашего слова в определенную устойчивую группу слов.

Ряд операторов позволяет согласовывать друг с другом различные части речи, находить в тексте любое нужное слово, отводить накопленные сведения в определенное место памяти и... наконец, давать слову перевод. Это конечное распоряжение после выяснения всех признаков.

Создание операторов очень упростило работу. И это в первую очередь почувствовали сами лингвисты. В своих схемах они просто указывали номер оператора в соответствии с действием, которое, по их мнению, должна была произвести машина. Такие схемы даже не надо было предварительно программировать.

Получив номер оператора, машина сразу включает его программу, в которой насчитывается около сотни команд, и выполняет нужное лингвисту действие. Один из сотрудников лаборатории как-то сравнил операторы со стандартными деталями, из которых можно сложить дом любой величины и архитектурного облика — от железнодорожной будки до восьмиэтажного здания.

Изучить операторы лингвисты могут очень просто. И таким образом они научатся машинному языку и станут непосредственно общаться с самой машиной. «Диспетчер» или «транслятор» (специальная программа, насчитывающая три с половиной тысячи команд) «поймет» эти операторы и согласует действие всех программ алгоритма. «Диспетчер» решает, что делать с данными, полученными в результате работы каждого оператора. Он координирует все операции на всех этапах работы алгоритма и, в случае неполадки, извещает людей.

Остается добавить, что алгоритм, созданный в лаборатории машинного перевода, годен для работы на быстродействующих электронных вычислительных машинах разных систем. Нужен лишь новый «диспетчер», приспособленный для новых условий работы. И этим

снимается обсуждавшаяся некоторыми специалистами идея создания специализированных переводческих машин.

Итак, машина закончила работу со словарем, узнав очень многое о всех словах текста, а именно, получив их цифровые эквиваленты. Эта информация позволит машине перейти к грамматическому анализу английского текста.



Считается, что английский язык относится к языкам аналитическим, а русский — к синтетическим. В аналитических языках глагольные формы чаще всего образуются при помощи вспомогательных глаголов, а вместо падежных окончаний употребляются предлоги. Английский язык имеет твердый порядок слов в предложении, а в русском языке, благодаря богатству форм слов, порядок свободный.

Лингвисту, занимающемуся машинным переводом, необходимо установить соответствия и различия структур языков. Надо знать, что в языках общее и в чем они отличаются друг от друга. Например, в английском у неодушевленных существительных нет рода, нет падежей, и тем не менее при переводе на русский существительные обретают род и падеж. Откуда же они берутся?

В алгоритме машинного перевода есть более ста схем, которые анализируют грамматически английский текст и поставляют сведения для образования русских слов и построения русской фразы.

Сначала машина делает как бы разбор предложения. Она определяет, каким членом предложения является каждое слово. Но машинная грамматика несколько отличается от школьной. Каждое слово входит в одну из семи групп. Проведя анализ, машина узнает, является ли слово подлежащим, сказуемым, дополнением, обстоятельством и т. д.

Поступает она точно так же, как школьник, разбирающий предложение у классной доски. Например, подлежащее она может определить по месту в предложении. Если машина «видит» слово «I» — «я», то она смело может сказать, что это подлежащее, так как иным членом предложения это слово может быть в одном случае на миллион. Перечисление всех признаков, по которым можно определить грамматическую группу слова, заняло бы у нас десятки страниц. (Кстати, в машинной грамматике «я» считается существительным, а прочие местоимения соответственно существительными или прилагательными, потому что они полностью подчиняются законам, по которым изменяются эти имена.)

Не зная еще русских слов, машина определит некоторые падежи. Подлежащее всегда стоит в именительном падеже. По так называемой саксонской форме она определяет родительный падеж. По подлежащему она узнает лицо стоящего рядом с ним глагола. По разным признакам она выясняет наклонение, время, число, степень сравнения, вид, залог и другие грамматические категории слов и все сведения аккуратно «записывает» в «ячейку» каждого введенного слова.

Очень важно разобраться в глаголе. Это центр синтаксической конструкции. И его анализ является основой для анализа фразы. При анализе глаголов постепенно накапливаются грамматические признаки для русского перевода. Если у глагола есть окончание «s», то мы с самого начала можем дать ему целую характеристику (3-е лицо, единственное число, сказуемое, настоящее время). Постепенно мы получаем о глаголе сведений все больше и больше.

Интересно отметить, что весь грамматический анализ машина ведет в пределах простого предложения. Но тот, кто изучал английский язык, знает, что в английских сложных предложениях запятые между простыми предложениями ставят редко и даже опускают слова «which»

и «that» («который» и «что»), так помогающие нам разбираться в больших фразах. Казалось бы, здесь мы зашли в тупик и окончательно запутаемся. Но лингвисты нашли выход и из этого положения. В машине предусмотрена программа, которая делит сложную фразу на простые предложения, расставляет запятые и даже восстанавливает пропущенные слова «который» и «что».

Но не все сведения для построения русской фразы можно получить из английской. Тут же нужен русский эквивалент английского слова. А у нас пока есть только его цифровой эквивалент. Из него мы, конечно, знаем многое. Например, слово «table». Мы знаем лишь, что это неодушевленное существительное. Оно многозначно и переводится на русский и как «стол», и как «таблица», и как «расписание». Только переведа, мы узнаем род слова. Значит, для дальнейшего грамматического анализа нам надо выяснить все возможные переводы этого слова.



И тут работа грамматической части машины прерывается, и начинает действовать ПМС (программа многозначных слов). Это одна из самых трудоемких частей алгоритма. Слов, имеющих по несколько значений, оказалась добрая четверть из записанных в словаре. Иные слова, как мы уже говорили раньше, имеют до сотни значений. И все эти значения надо определить по контексту, но теперь уже опираясь на обильную грамматическую информацию.

Более того, перевод слова всегда диктовался желанием сделать так, чтобы вся фраза звучала по-русски. И слово получало все новые и новые эквиваленты. Типично английские выражения переводились типично русскими.

Давайте вернемся к английской фразе «Last night I went to bed at ten o'clock». Вспомним, что сочли правиль-

ным следующий ее перевод: «Вчера вечером я лег спать в десять часов». А ведь если перевести буквально, то у нас получится: «Прошлой ночью я пошел в постель в десять часов». Казалось бы, на этом можно успокоиться. Но нет. Во-первых, по схеме омографов мы заменили «прошлой ночью» на «вчера вечером», а по программе многозначных слов «went» («пошел»), кроме многочисленных других переводов, из стилистических соображений получит еще и перевод «лег». Следующие два слова «to bed» тут же получают перевод «спать» и будут исключены из дальнейшего анализа. И это очень удобно. -И машине меньше работы и перевод правильный.

Всякому сведущему в переводческом деле человеку в первую очередь приходит в голову мысль, что машина-переводчик, опирающаяся в основном на формальные признаки, должна грешить буквализмом.

Но она не буквалистка, нет. И это один из парадоксов машинного перевода.

В будущем, при очень объемной памяти, она сможет запомнить все устойчивые выражения, идиомы и наиболее часто употребляющиеся фразы одного языка и сразу передавать их смысл выражениями и идиомами другого языка. Их, правда, будут не тысячи, а десятки или даже сотни тысяч. Но уже сейчас машина, встретив фразу «How do you do!», переведет ее как «Здравствуйте!».

Узнав номер каждого русского слова, которым мы переводим английское, можно теперь получить и остальные грамматические сведения. На этом этапе мы узнаем не только род слова, но и падеж его.

И опять мы опираемся на ближайшее слово. На предлог, например. Мы говорим «на столе», но «под столом». В этом примере закономерность видна невооруженным глазом. Но надо было проделать большую работу по наблюдению над языком, прежде чем выявились все закономерности. А выяснив грамматическую характеристику существительного, мы можем перенести полученные

сведения и на прилагательное, которое в русском языке обычно согласуется с существительным в роде, числе и падеже, а в английском просто примыкает к нему. Теперь, когда мы уже знаем перевод английских слов и их грамматические характеристики, казалось бы, надо приступить к построению русской фразы и дать ее прочесть нетерпеливым зрителям, собравшимся у буквопечатающего устройства машины. Но, нет. Машине надо еще отредактировать эту фразу. Странно, скажете вы, редактировать фразу, которой еще нет.

Да, таков еще один из парадоксов машинного перевода.

●

У английского языка свои законы, а у русского — свои. У английского языка один порядок слов, а у русского — другой. В английской фразе никогда не может быть двух отрицаний, а в русской мы только что употребили их два: «никогда» и «не». Английская фраза в буквальном переводе звучала бы так: «В английской фразе никогда может быть двух отрицаний». Англичане обходятся одним «никогда», но нам, чтобы машина выражалась по-русски, пришлось наделить ее программой, которая вставляла в английскую фразу еще и «не».

Русский язык очень гибок, и он позволяет сохранять английский порядок слов во фразе. Но не всегда. Английская фраза «He was not ready» буквально переводится как «Он был не готов». Такой порядок слов режет слух, и мы меняем английский порядок слов на «Он не был готов».

Я привел примеры самые простые, но программа редактирования, реконструкции и перестановки решает задачи и очень сложные, связанные с коренной переработкой английской фразы.

Итак, английские слова расставлены в должном (русском) порядке, и, наконец, машина может, используя

грамматическую информацию, накопленную при каждом слове, писать по-русски.

Этому ее научили специалисты в области русского языка во главе с Еленой Федоровной Калининой. Она пришла работать в лабораторию машинного перевода в 1958 году, когда еще только составлялись словарь и схемы многозначных слов. Я помню ее первые страхи и колебания, когда она стала понимать всю огромность дела, которое ей предстояло осуществить. Никаких руководств и печатных трудов по русской части алгоритма машинного перевода тогда еще не было.

За рассуждениями лингвистов об эмоциональных ослужнениях и смысловых оттенках в языке надо было разглядеть конкретные признаки, которыми выражались все эти сложности. Чтение работ больших русских ученых побуждало к пристальному наблюдению за родным языком.

«Я чувствовала себя,— говорила впоследствии Калинина,— как человек, не умеющий плавать и брошенный в воду». Это чувство было знакомо всем нам, взявшимся за создание алгоритма машинного перевода.

Научить машину говорить по-русски? Мы пишем: «Я вижу дом», но «Я вижу человека». Машина должна ощущать элемент одушевленности и соответственно наделить слова окончаниями.

Калинина начала думать, каким должен быть русский словарь. В свое время Бельская помещала в словарь существительные в именительном падеже. Калинина пришла к выводу, что более правильно брать основы слов. И не одну основу каждого слова, а несколько.

И тут начинается расхождение с традиционной грамматикой. В машинной грамматике потребовалось выделить у существительных две основы, у глаголов — три, а у прилагательных целых четыре.

Вернемся к нашему примеру «Last night I went to bed...». Глагол «went», как вы выяснили, переводится словом

«лег» («лечь»). И еще нам известно по форме глагола, что он прошедшего времени. В машине даны основы слова «лечь». Они суть «ле», «ляг» и «леж». От них можно образовать любую форму слова — глагол во всех временах, причастие, деепричастие... По специальной таблице машина выясняет, что прошедшее время образуется от основы «ле». В грамматической информации при этом слове значится, что глагол должен быть первого лица, единственного числа и, предположим, мужского рода. По таблице значится, что при таком сочетании признаков надо присоединить к основе «г». И у нас получается: «Вчера вечером я лег...»

В схемах русской части алгоритма указано, как образуются причастия, деепричастия, возвратные глаголы, степени сравнения прилагательных и т. д.

Особенно трудно было разрешить вопрос падежа. Из английских грамматических схем не всегда можно почерпнуть сведения о падеже.

Были составлены списки глаголов, управляющих определенными падежами существительных. (Например, глагол «руководить» требует, чтобы после него существительное стояло в творительном падеже — «чем?».) Все русские слова разбиты на лексические группы, которые обуславливают определенные грамматические формы. А всего таких лексических групп семьдесят семь.

Начинающие изучать иностранный язык, переводя какой-нибудь текст, любят говорить: «Вот здесь я догадался по смыслу». На зачисление слова в ту или иную лексическую группу влияет главным образом смысл слова. И здесь уже можно было бы говорить об «осмысленной» работе машины, если бы до нее смысловой анализ многих тысяч слов не провели сами люди и не дали ей эту осмысленность в готовом виде, в форме удобных для пользования таблиц.

Беглое описание алгоритма машинного перевода завершено. В таком алгоритме заранее предусмотрены

все логические операции, и фраза попадает в машину как на конвейер, у которого нет лишних движений.

Однако, если мы хотим увеличить «интеллект» машины, нам необходимо предусмотреть в ее работе и другие особенности человеческого мышления. Человек сам создает для себя алгоритмы и этим облегчает себе работу. Составив правила перевода, человек следует им. А нельзя ли и машине составлять для самой себя такие правила? Например, выявлять категории языка, определять одушевленность и неодушевленность существительных? В Киевском институте кибернетики проделали такой опыт. В машину ввели около двухсот существительных и глаголов, не указывая их категорий.

Затем составили на основе этого словаря ряд фраз. Одни из них имели смысл, другие были бессмысленны. «Женщина варит обед, инженер делает проект, станок делает гайки, лодка варит обед». Причем всякий раз вводилась информация, имеет ли фраза смысл или нет. По программе машина разбила слова на несколько групп, выделив категорию людей. Проанализировав фразы: «Женщина варит обед» (есть смысл) и «Лодка варит обед» (нет смысла), она «запомнила», что со словом «варит» слово «женщина» сочетаться может. Затем машине поручили самой определять, какие фразы имеют смысл. Для проверки в машину ввели фразу: «Инженер варит обед», и машина ответила, что это бессмыслица. Из анализа предложений «Инженер делает проект» и «Станок делает гайки» она не могла получить нужной информации. Она выделила «инженера» в категорию людей только после того, как встретила это слово с глаголами, сочетающимися только с одушевленными существительными.

В этом опыте уже моделировались некоторые связи, напоминающие ассоциации, которые возникают у человека.

Американские исследователи опробовали свою систе-

му на 400 элементарных английских предложениях, и она сама вывела грамматику, по которой при вторичном прохождении правильно проанализировала все 400 предложений. В дальнейшем ей дали поучиться на трехстах предложениях, после чего она правильно разобрала три четверти из ста новых предложений. Но пока машина имела дело лишь с упрощенным английским языком...

●
Наш алгоритм в принципе работает. Может быть, ему еще не хватает совершенства. Но что совершенно на свете?

Полное удовлетворение сделанным недостижимо, потому что каждый прожитый день выдвигает новые требования, и совершенствовать машинный перевод теперь придется из года в год, так как необходимость в нем будет расти.

Уже не воспринимается как мечта составление схем многозначных слов на самой машине. Когда-то это потребовало нескольких лет работы большого коллектива лингвистов. Сегодня делаются наброски программы, которая позволит машине самой анализировать слово и его окружение, выбирать признаки и давать перевод.

Уже машина учится выбирать слова и пополнять ими собственный словарь. Задав некоторую информацию, можно заставить машину саму совершенствовать таблицы анализа русской части.

Машина научилась русскому языку, но все правила ей преподносили в готовом виде. Сейчас машина учится составлять для себя правила — определять род слова, его основу и т. д. Машинный алгоритм можно использовать даже для обучения языку.

Но не будем говорить о том, что еще не завершено, а вернемся к испытаниям алгоритма.

В машинный зал приходят человек двадцать. Пятеро работают, остальные любопытствуют, сочувствуют и со-

здают сутолоку. В машину вводятся сначала все программы алгоритма, а затем английский текст. Перевод начался.

Но не все идет гладко. Есть еще много «огрехов» в программах, и ошибки сразу путают все карты. Вот здесь программист неправильно указал адрес—«ячейку», и М|шина, как говорят, «зациклилась». Она вновь и вновь повторяет одну и ту же операцию и похожа на патефон, игла которого бесконечно ходит по одной борозде пластинки. Надо исправлять программу.

Бывает, что виновата и сама машина. В ней тысячи и тысячи деталей. Выход из строя хотя бы одной может вызвать появление разряда-паразита. И вдруг машина вырабатывает команду — стереть все из своей памяти. Погибает весь цикл перевода. Поэтому инженеры так борются за надежность работы машины.

Для избежания подобных аварий разработана специальная программа — система блокировки. Она проверяет правильность работы машины и следит, как вырабатывается информация на каждом этапе алгоритма. Эта система указывает, что испортилось и где. Сбой ликвидируется, и работа машины продолжается или останавливается, чтобы не стерлась полученная информация.

По специальным критериям система даже проверяет правильность информации, и если подлежащее вдруг окажется в дательном падеже («Ему пришел в гости»), то об этом будет немедленно сообщено людям. На выработку и совершенствование такой системы ушел целый год работы нескольких специалистов.

Застучал телетайп. Первые строки. Правильно, правильно... а здесь ошибка. Не обходится и без курьезов. При переводе статьи из свежей газеты машина перевела фамилию англичанина «Vowels», как «кишки». Она не так уж ошиблась, перевела правильно. Только здесь переводить не надо. Следовало написать фамилию Бауэлс, и все. Ей бы еще немного чувства юмора...

Ошибки можно исправить. Алгоритм еще надо «доводить», совершенствовать. Главное — машина может переводить. Она может заменить переводчика.

Машина не в состоянии переводить любой текст. Но она уже справляется с газетой, с техническим журналом... Она подает большие надежды. Однако художественным переводом заниматься ей не под силу, потому что творческие возможности машины весьма ограничены...



лектронные вычислительные машины становились верными помощниками человека, взяв на себя некоторые несложные функции управления производством, счетное дело, перевод с одного языка на другой... Помогли они и в литературных изысканиях.

Так, один машинный эксперимент дал возможность разрешить вековой спор, развернувшийся вокруг литературной проблемы.

Автором двух самых известных эпических поэм Древней Греции считается Гомер.

Тысячи лет «Илиада» и «Одиссея» вдохновляли поэтов и были мучением для учеников, которых заставляли заучивать тяжелые гекзаметры.

«Илиада» повествует об осаде греками малоазиатского города Трои, о спорах богов, сонм которых возглавлял всемогущий Зевс, о гневе жестокого героя Ахилла, обидевшегося на царя греков Агамемнона.

Существуют девять жизнеописаний Гомера, но все они написаны уже во времена римских императоров. По преданию, Гомер был слепым. Честь считаться родиной Гомера оспаривали семь греческих городов: Кума, Смирна, Хиос, Колофон, Пафос, Аргос, Афины. Одни говорили, что он жил во времена Троянской войны, другие — позже.

Но вот в XVIII веке немецкий филолог Август Вольф, анализируя поэмы, пришел к выводу, что «Илиада» и

«Одиссея» не могли принадлежать одному поэту, а являлись плодом творчества многих поэтов и певцов. Он утверждал, что отдельные песни были объединены в поэмы через много столетий после того, как они впервые исполнялись, и что отредактировали поэмы люди весьма посредственных умственных способностей.

По Вольфу, такого поэта, как Гомер, никогда не существовало. Немецкий ученый опирался на серьезные факты. Еще в Древней Греции многие считали, что Гомер совсем не историческая личность, а условное обозначение для целого коллектива поэтов. Да и в наше время академик Н. Я. Марр утверждал, что слово «гомер», которое у басков и грузин значит «слепец», — синоним собирателя, рапсода, поэта, обрабатывавшего и исполнявшего народные песни.

Так возникла одна из самых запутанных научных проблем, которую не могли разрешить сотни лет и называли «гомеровским вопросом».

«Вольфианцам» противостояли «унитаристы» — сторонники индивидуального творчества. Если первые находили в поэмах массу несуразностей, то вторые указывали на то, что такие же несуразности можно найти и в произведениях современных писателей. Так, женой Гефеста в «Илиаде» была Харита, а в «Одиссее» — Афродита, у Нестора в «Илиаде» было одиннадцать братьев, а в «Одиссее» только два, и т. д. «Унитаристы» приводили примеры подобных же несуразиц из Гёте и Толстого. И те и другие за полтора века написали тысячи статей и книг, посвященных гомеровскому вопросу.

Все эти бесконечные споры были разрешены при помощи электронной вычислительной машины. Американский филолог Джеймс Макдонуг заставил ее проанализировать все 15 693 строки «Илиады». Машина сопоставила метрику стиха разных частей поэмы, выявила некоторые общие для всего произведения стилистические

особенности. Статистические подсчеты и анализ многочисленных примеров, приведенных машиной, позволили филологу доказательно утверждать, что поэма написана одним человеком и что Гомер — личность историческая.

В Советском Союзе новосибирские ученые использовали электронную быстродействующую машину для расшифровки письменности древних майя, о чем я в свое время рассказал в книге «Загадочные письмена».



Будущее сулило гигантское развитие способностей машин, которые проявляли чудеса оперативности в решении некоторых логических задач. И вот тогда была выдвинута идея, что в принципе возможно создание живых и мыслящих существ, способных к самостоятельному творчеству.

Сперва, захлебываясь, говорили о возможности воссоздания творческой личности, потом наступило отрезвление. Творцу (писателю, например) нужны воспоминания о восприятии природы, о собственных радостях и невзгодах в детстве, о муках болезни роста, о событиях, менявших течение жизни многих людей, о чувствах, испытанных к людям... Создать модель какого-нибудь талантливого поэта значит промоделировать все, что он видел в жизни, все общество, в котором он жил, природу, которую он видел, всех его предков, историю общества, Вселенную...

Федор Михайлович Достоевский, пытавшийся разобрататься в сущности творчества, писал:

«...Поэма, по-моему, является как самородный драгоценный камень, алмаз, в душе поэта, совсем готовый, во всей своей сущности, и вот это первое дело поэта, как *создателя и творца*, первая часть его творения... Затем уж следует *второе* дело поэта, уже не так глубокое и таинственное, а только как художника: это, получив

алмаз, обделать и оправить его. Тут поэт — почти только что ювелир».

Перевод художественных произведений следует отнести ко «второму делу», которое тоже делает художник, но не творец, а опытный, понимающий чужое творчество и талантливый ремесленник. Это понимание чужого творчества и предусмотрено оговоркой Достоевского («почти только что ювелир»), оно требует жизненного, человеческого опыта, который лишь отчасти дает образование и начитанность.

Наделить такими качествами машину, разумеется, невозможно.



Как работает человеческий мозг? Ответа на этот вопрос ищут и психологи, и биохимики, и нейрофизиологи, и анатомы.

Мозг — загадочная материя, над тайной которой задумывались еще мудрецы Древней Греции. Великий врач средневековья Везалий первым сказал, что «мозг построен ради главенства разума, а также чувствительности и движения, зависящего от нашей воли». Декарт открыл рефлексы, а Павлов — сигнальные системы мозга. В конце XIX века нашли, что каждый участок мозга выполняет определенные функции. Сейчас усиленно анализируются биотоки мозга. Кибернетика стремится строить модели, выполняющие отдельные функции его. Однако науку о мозге по-прежнему можно отнести к ряду описательных, а не точных наук.

Без разума величие Вселенной было бы неполным. Материя, способная рождать мысль, весит в среднем **1300** граммов. Умственные способности человека не зависят от ее веса. Мозг Тургенева весил 2012 г, Голсуорси — **1482** г, а Франса — **1017** г, Бехтерева — 1720 г, а Гаусса — 1492 г. Самый тяжелый мозг оказался у одного слабоумного (2850 г).

На питание мозговых тканей расходуется 1/5 всей крови человека.

Мозг построен из серого (клетки) и белого (волокна) вещества и состоит из многих причудливо переходящих друг в друга образований. Приходящая информация анализируется в коре — сером веществе, покрывающем тонким слоем полушария мозга. Это она состоит из миллиардов нервных клеток, размеры которых колеблются от 5 до 200 микронов. Они способны действовать безотказно в течение многих лет. Каждая клетка имеет 3—4 тысячи контактов, а для всего головного мозга число связей достигает поистине астрономической цифры — шестидесяти тысяч миллиардов.

Мы можем и дальше продолжать наше описание, но оно ровно ничего не скажет о процессе рождения мысли. Потому-то ученые, описывая предположительную работу мозга, чаще всего прибегают к аналогиям.

Американец Маккаллох характеризовал мозг человека как трехфунтовую вычислительную машину мощностью в 25 ватт с емкостью накопителя в 13—15 миллиардов битов, обладающую активным полупериодом в течение полусуток и регенерацией пяти процентов содержания, накопленного за 50 лет.

Остроумно, но не боле...

Американский психолог А. Чапанис предложил такую схему специфических особенностей работы человека и машины:

Человек

Машина

1. Способен решать альтернативы малой вероятности, т. е. отвечать на неожиданные события

1. Отличается большой сложностью программирования, так как трудно предвидеть все возмож-

2. Способен использовать восприятие и пространство и превращать обрывки информации в значительное взаимосвязанное целое.

3. Знает различные способы выполнения действий. Может выполнять те же или подобные действия и достигать тех же результатов другими способами в случае, если основные способы применить невозможно.

4. Обладает ограниченной способностью, как канал связи, т. е. максимальное количество информации, обрабатываемое им за единицу времени, невелико.

ные события и составить программу, предусматривающую все варианты.

2. Характеризуется нулевой или очень ограниченной способностью «осознавать». Структура работы машины должна быть тщательно запрограммирована, причем существует много различных способов компоновки структуры из различных элементов.

3. Альтернативные способы выполнения действий ограничены. Может совершенно выйти из строя при частичной поломке.

4. Как канал связи, позволяет добиться любой желаемой пропускной способности.

5. Ввиду утомляемости и рассеивания внимания снижает свою работоспособность в сравнительно короткое время.

6. Производит вычислительные операции сравнительно медленно и неточно.

5. Ухудшает свое «поведение» лишь спустя значительный промежуток времени.

6. Производит вычислительные операции исключительно точно и очень быстро.

Кроме того, у человека есть еще одна способность, которая позволяет ему, минуя логический путь, открывать истины непосредственно, как бы видеть их «в озарении» («первое дело» по Достоевскому). Это интуиция.

Еще в XVII веке Декарт считал, что аксиомы, лежащие в основе науки, открыты интуитивно. А уж потом при помощи логических рассуждений вывели из них все остальные знания. В XX веке великий математик Анри Пуанкаре бросил крылатую фразу: «Посредством логики доказывают, посредством интуиции изобретают». Он же пояснил взаимоотношение логики и интуиции: «Логика говорит нам, что на таком-то и на таком-то пути мы, наверное, не встретим препятствий; но она не говорит, каков путь, который ведет к цели. Для этого надо издали видеть цель, а способность, научающая нас видеть, есть интуиция. Без нее геометр был бы похож на того писателя, который безупречен в правописании, но у которого нет мыслей».

Было много споров по этому поводу, но и теперь психологи говорят, что «в настоящее время проблема логики и интуиции в математике остается открытой».

А нужна ли интуиция переводчику? Безусловно. Разумеется, не для поисков аксиом. Но в переводе есть свои маленькие открытия, которых не сделать без вдохновения. Например, передать непере译димую и не объясненную словарями идиому так образно и емко, что она войдет в золотой фонд родного языка. Таких случаев было немало. Иногда это делалось интуитивно.

Предусмотреть нечто подобное при составлении алгоритма машинного перевода невозможно. С другой стороны, переводчик-человек обладает «способностью», даже не разобравшись до конца в тексте или не найдя подходящего эквивалента в словаре, перевести непонятное место туманно, но все-таки складно. Это уже халтура, искажение информации, на которое машина «не пойдет» — просто не даст перевода. Кстати, в истории человечества открывалось интуитивно немало «аксиом», из которых логично выводили целые учения, оказывавшиеся в конце концов ложными...



Во Всесоюзной патентно-технической библиотеке хранятся миллионы описаний изобретений, еще не переведенных на русский язык. Причина все та же. Не хватает переводчиков. А ведь патентные тексты составляются так, чтобы в них не было «нюансов и подтекстов». Это термины и четкие формулировки. Кому, как не машине, заняться их переводом. Более десяти лет разрабатывались специализированные алгоритмы для перевода патентной информации. В 1969 году в электронную вычислительную машину «Раздан-3» введен комплекс из 16 основных и вспомогательных программ — 20 тысяч машинных команд. Машина тратит на перевод с английского одной публикации (150—300 слов) от двух до пяти минут. Скорость перевода пока еще невелика, но на более совершенной машине она будет выше. Сейчас отдел автоматизации перевода и

индексирования работает над алгоритмом для перевода японских патентов.

Создана машина, которая переводит с английского информационные тексты. Только русский автоматический словарь ее включает около сорока тысяч лексических единиц.

Вопросами машинного перевода занимаются в Ленинграде, Кишиневе, Минске, Махач-Кале, Алма-Ате...

Главное сейчас — добиться высокой рентабельности перевода. Пока еще велико машинное время, но прогресс в этой области уже виден.

Много систем машинного перевода работает за рубежом. Одно перечисление их заняло бы у нас не менее страницы. Расскажем об одной из них.

На недавней конференции по вопросам американо-советских торговых отношений, которая была организована в Вашингтоне Национальной ассоциацией промышленников, участники увидели в вестибюле отеля «Шорем» входное устройство машины «Логос-1». Его установила компания «Логос девелопмент», снабдив письменной рекомендацией: если вы хотите торговать с Советским Союзом, вам придется рассказывать о своих товарах по-русски.

Желающие диктовали оператору английские фразы технического порядка, тот нажимал клавиши телетайпа, сигналы предавались по проводам в электронную машину, находившуюся в одном из городков штата Нью-Йорк. Не проходило и минуты, как аппарат выдавал перевод фразы на русский. Фраза иной раз получалась корявая, но технически грамотная. Например: «Регулирование выхода газов печей для выдержки жидкого металла должно быть осуществлено посредством вытяжного зонта бортовой тяги и четвертого отверстий». Требовалась правка, но для сведения сошло бы и так...

В рекламном проспекте фирмы говорится, что необходимость создания машины была вызвана «расширя-

ющеися торговлей между Востоком и Западом, невероятно возросшей нуждой в переводах с английского на русский, чешский, румынский и другие языки». Ученые, мол, признали, что с их техническим словарем легче управляться, чем с нюансами художественной прозы. «Поэтому они совершенно отказались от мысли создавать машины для перевода Шекспира на китайский и стали программировать компьютеры для перевода только технической и коммерческой прозы». К рекламе приложена карта СССР.

Алгоритм перевода был создан работниками компании во главе с Бернардом Е. Скоттом. Намечается, что память машины в ближайшее время вберет в себя словарь на 300 тысяч слов. Если учесть, что в советской машине словник дополняется словарем словосочетаний, то емкость памяти советской и американской машин одинакова. Скорость перевода у американской машины выше. Но наша машина переводит более тщательно, редактировать ее переводы не требуется.

Фирма, создавшая этот алгоритм, уже принимает заказы на переводы. Ее услугами воспользовалась компания «Суинделл-Дрекслер», участвующая в разработке технического проекта строительства КамАЗа. Текст для перевода на машину может передаваться из любого города США. Фирма предложила советским представителям сотрудничество в области машинного перевода...



Когда произведения писателя начинают пародировать, то это равносильно признанию его таланта. Многочисленные шутки в адрес вычислительных машин говорят об их упрочившемся положении в нашем мире. Рассказывают, что некая машина просто отпечатала: «Cogito, ergo sum» («Я мыслю, следовательно, существую»). Есть анекдот о том, как в Ватикане в элек-

тронную машину ввели сочинения святого Фомы Аквинского и спросили: «Есть ли бог?» Обработав полученную информацию, машина ответила: «Теперь есть». Известна карикатура, на которой ученый с изумлением смотрит на большую электронную машину и говорит: «Пяти тысячам математиков понадобилось бы четыре тысячи лет, чтобы сделать такую ошибку!»

Может возникнуть предположение, что с развитием машинного перевода переводчики-люди останутся без работы. В некоторых докладах о механизации лингвистических работ в шутку цитировались стихи В. Брюсова: «Но вас, кто меня уничтожит, встречаю приветственным гимном». На это можно ответить рассказом об одном любопытном факте.

Когда было налажено производство электронных вычислительных машин, поговаривали, что без работы останутся математики-вычислители. Машины теперь выполняют работу миллионов математиков, но только на одно обслуживание машин потребовалось больше математиков, чем их прежде было вообще. Просто наука движется семимильными шагами, объем работ неслыханно увеличивается и требует все больше людей и машин. То же произойдет и с переводчиками. Художественная литература будет переводиться людьми всегда. А о росте объема научной и технической информации мы уже говорили.

Машинный перевод воплощает в себе опыт многих переводчиков. Пройдет еще сравнительно немного времени, и машина будет «знать» гораздо больше любого переводчика технических и научных текстов, взятого отдельно. Главное — передать машине все, что знаем мы.

И машина воздаст сторицей. В процессе «обучения» машин языку мы, люди, сами лучше познаем его.

Но уже сейчас кое-кто склонен говорить о себе и о машине, как о коллегах:

«Мы — переводчики».