

ОПЫТЫ МАШИННОГО ПЕРЕВОДА

К ВОПРОСУ О ВОЗМОЖНОСТЯХ МАШИННОГО ПЕРЕВОДА

В связи с созданием сложной электронной вычислительной машины в последнее время оживленно обсуждаются практические и теоретические вопросы кибернетики, в частности, вопрос о выполнении машинным путем перевода текста с одного языка на другой. При тесном сотрудничестве лингвистов, математиков и инженеров уже осуществлен опыт такого перевода¹. Оценивая результаты этого опыта с точки зрения лингвистической проблематики, возможно и необходимо рассмотреть перспективы развития машинного перевода.

В техническом отношении совершенствование специальной электронной машины для перевода идет быстрым темпом — выдвигаются проекты построения быстродействующей машины (30 тыс. и более тактов в секунду, т. е. 10 и более предложений) с применением новейших достижений техники (полупроводники, ферриты, устройство долговременной памяти). Эта машина в будущем даст возможность обслуживать даже своеобразные телефонные абоненты (одна машина на 1200 абонентов) или выполнять синхронно устный перевод, например на конференциях². Но так как машинный перевод представляет собой не только техническую, но в значительной мере и лингвистическую проблему, выяснение возможностей лингвистики для осуществления механизированного перевода имеет принципиальное значение.

Как известно, работа машины по переводу языкового текста предполагает построение специальной программы, которая вводится в закодированном виде в машину и на основании которой машина совершает операции по переводу. Так как машина может функционировать на основе этой программы только по принципу включения и выключения контакта (положительный или отрицательный ответ), то программа должна содержать описание последовательности всех заданных однозначных операций для получения искомого результата³.

В конечном итоге система всех операций должна основываться на определенном алгоритме, в котором лингвистические единицы приравнены к однозначным символическим величинам, ибо машина может оперировать только с постоянными, формально зафиксированными элементами. Отклонение от этой программы в принципе не допускается, если считать, что и корректировка программы в свою очередь может быть основана только на соответствующей программе. В такой же мере программирование, выполняемое самой машиной, ее «самоастройка» могут быть произведены в конечном итоге только на основе заданной человеком определенной команды. Составление подобной программы предполагает поэтому полную формализацию системы языка, т. е. приведение его структуры — как лексической, так и грамматической — к символическому уравнению, которое может служить целям однозначных действий электронной машины (действие ее в данном случае аналогично процессу математического вычисления машинным путем)⁴.

Специалисты прямо утверждают, что «... механизация некоторых функций возможна тогда, когда они в принципе описываются конечным алгоритмом и когда мы настолько детально изучили процесс выполнения этих функций, что умеем его расчленивать на последовательные акты, изображаемые в машине элементарными логическими

¹ См., например: «Machine translation of languages», ed. by W. N. Lock e and A. D. Booth, New York — London, 1955; Д. Ю. Панов, Автоматический перевод, М., 1956; И. К. Бельская и др., Некоторые вопросы автоматизации перевода, «Вестник АН СССР», 1956, № 12.

² См. Л. И. Гутенмахер, Статистические и информационные машины нового типа, «Вестник АН СССР», 1956, № 10.

³ См., например: О. С. Кулагина и И. А. Мельчук, Машинный перевод с французского языка на русский, ВЯ, 1956, № 5; И. К. Бельская и др., указ. статья, стр. 28—29.

⁴ А. И. Китов, Электронные цифровые машины, М., 1956.

операциями»¹. Но так как условием функционирования машины для выполнения ею специальной задачи — именно перевода — является наличие формальной программы, то решение этой задачи целиком выпадает на долю лингвистической науки. Здесь языкознание оказывается перед важной теоретической проблемой — исследованием возможностей и границ формализации языка.

Проведенные в СССР и США опыты перевода с заданного языка (английского, французского) на русский и наоборот были основаны на сознательном ограничении характера текста (узко технического), его словаря (до 2 тыс. единиц), фразеологии (отсутствие сочетаний с переносным значением, идиом и т. п.), многозначности слова и других языковых явлений. Составление программы пока что было возможно только при выполнении этих условий. Совершенно естественно, что формализация правил для лингвистических соответствий в таких узких рамках могла быть, после определенного предварительного исследования, уложена в соответствующую программу — команду для машины. Несмотря на большое количество операций, потребовавшихся для получения конечного результата, при существующей скорости (8—16 тыс. операций в сек.) перевод был осуществлен сравнительно быстро (один абзац в 8—10 фраз за 18—20 сек.). В принципе эта работа могла быть осуществлена и каким-либо другим механическим способом, так как все операции по переводу, зафиксированные в программе, расчленимы на простейшие акты, моделирующие элементарные последовательные процессы работы переводчика по отысканию требуемых эквивалентов. При колоссальной скорости работы электронной машины эти элементарные акты исполнялись мгновенно, благодаря чему при переводе не только быстро отыскивались готовые лексические элементы, но и учитывались соответствующие грамматические форманты (например, падежные формы, порядок слов, некоторые глагольные формы), в обобщенном виде представленные в программе-команде.

Перевод, сделанный для этих простых текстов, был удовлетворителен (в частности, на машине БЭСМ АН СССР (англ. яз.) и на машине «Стрела» (франц. яз.)) и требовал незначительной работы редактора². На основании этих опытов теперь можно утверждать, что удовлетворительный перевод с одного языка на другой при четко ограниченных лингвистических условиях может быть осуществлен машинным путем.

*

Как же далеко могут быть раздвинуты границы применения машинного перевода, какие существуют в данном случае сдерживающие факторы, связаны ли эти ограничения с техническим прогрессом или они обусловлены существом самого предмета, можно ли ожидать в будущем решения вопроса о возможности перевода текста любого характера — литературного, публицистического, поэтического и т. д. и с любого языка — вот вопросы, которые должны быть принципиально исследованы теоретическим путем. Естественно, что исходным моментом в решении данных вопросов должно быть исследование сущности языковой системы как базы, на которой строится сам процесс перевода.

Некоторые авторы полагают, что коммуникативная функция языка как раз и предопределяет возможность безграничной формализации языка, его способности в этом отношении быть представленным в необходимой символической программе «в удобной для машинного перевода форме»³.

Коммуникативную функцию языка эти авторы рассматривают, видимо, как относительно самостоятельную функцию сообщения какому-либо лицу какого-либо содержания, и с этой точки зрения они обосновывают возможность изобразить формальными средствами коммуникативное ядро любого текста и подготовить его таким образом для машинного перевода. Нам кажется, однако, что неправомерно отгораживать в какой-либо степени так называемую коммуникативную функцию языка от других его функций (например, экспрессивной), так как язык представляет собою единую сущность и систему, в которой неразрывно соединены все функции, определяющие его общественную природу. В живом языке экспрессивная, коммуникативная и все другие функции языка взаимно обусловлены и не делимы друг от друга; существование одной функции предполагает наличие другой. Выделять и зафиксировать каким-либо образом, в каком-либо живом языковом акте одну из сторон принципиально невозможно, поэтому вряд ли будет оправдана попытка найти способ формализовать коммуникативную сторону языка для целей машинного перевода.

¹ М. В. Келдыш, А. А. Ляпунов, М. Р. Шура-Бура, Математические вопросы теории счетных машин, «Вестник АН СССР», 1956, № 11, стр. 35.

² См.: И. К. Бельская и др., указ. статья, стр. 31; И. А. Мельчук, Сопоставление по вопросам разработки и построения информационных машин, ВЯ, 1957, № 5, стр. 161—162.

³ П. С. Кузнецов, А. А. Ляпунов, А. А. Реформатский, Основные проблемы машинного перевода, ВЯ, 1956, № 5, стр. 108.

В любом цельном языковом произведении нельзя «отпрепарировать» какую-либо одну сторону языка, не исказив его действительной природы. В предполагаемой форме может существовать только искусственный язык (например, эсперанто), но он уже не нуждается ни в каком переводе. Создание искусственного языка не облегчает нашей лингвистической задачи, а является лишь ее другим решением, которое относится уже к иной области. Это не перевод при наличии множества языков, а замена множества языков одним языком.

Другой путь, предлагаемый для принципиального решения проблемы машинного перевода, — путь, снимающий все трудности, связанные с формализацией языка, состоит, по мнению других авторов (например, логика Айдукевича, В. Уивера ¹), в том, чтобы создать некий «универсальный» язык, поддающийся полной алгоритмизации. Подобное мероприятие, во-первых, трудно в настоящее время обсуждать, поскольку нет еще работ, содержащих конкретное описание подобного языка, во-вторых, оно не связано с задачей непосредственного перевода с одного языка на другой, а предполагает создание «промежуточного» языка.

Создание некоего «метаязыка» в качестве посредника при машинном переводе с одного языка на другой имеет действительно важное практическое и теоретическое значение, но сам по себе «метаязык» в этом случае останется лишь вспомогательным промежуточным звеном, которое не снимает взаимодействия языков при окончательном переводе с одного конкретного языка на другой. В конечном итоге лингвистическая предпосылка перевода остается неизменной.

Возможный путь решения проблемы универсального машинного перевода (включая даже стихотворный перевод поэтических текстов) некоторые лингвисты видят, по существу, только в прогрессе техники ²; но так как в данном случае остается недоказанной возможность полной формализации конкретного языка, то эти прогнозы мало убеждают.

Довольно широко распространена еще одна точка зрения на перспективы машинного перевода ³. Она усматривает выход из затруднения в работе редактора машинного перевода и в соответственном расширении его функций. Практически такое решение вопроса всегда возможно, но оно не только ограничивает значение машинного перевода, но и ставит под сомнение его фактическую целесообразность в широких масштабах. Кроме того, здесь еще далеко не определены функции и объем работы редактора.

Все приведенные выше факторы еще раз убеждают нас в том, что проблема машинного перевода в настоящее время принципиально остается лингвистической проблемой и решение ее необходимо искать только в этой области.

*

Лингвистическая наука не может дать в настоящее время удовлетворительного однозначного решения при составлении формальных правил употребления многих языковых единиц. Такое состояние науки, однако, объясняется, по нашему мнению, не беспомощностью ее методики исследования — корни этого лежат в самой природе языка, исключающей возможность универсального формально-схематического описания функционирования элементов живого языка как единой системы.

«Правила перевода, пригодные для машины, не могут апеллировать к смыслу фразы, а должны исходить только из формы ее написания» ⁴, — это предпосылка работы машины. Возникает вопрос: возможен ли вообще перевод без понимания смысла переводимого? Смысл, семантика есть не что-то внешнее для языка, а неотъемлемая составная часть всех языковых единиц. Все элементы языка, все его формы объединены в систему разнообразными семантическими связями, которые нельзя уложить в единую, практически конечную схему — таблицу.

Рассмотрим основные причины этого явления. Известно, что большинство слов в языке имеет не одно, а несколько значений. Многозначность слова — его неотъемлемое качество. Решение вопроса о выборе соответствующего значения слова принципиально не связано с учетом его формальных признаков, а производится в процессе перевода по смысловым, семантическим признакам. Лингвистическая наука не в состоянии указать какие-либо отправные данные, необходимые в этих случаях для составления программы машинного перевода. В словарях зафиксированы десятки, даже сотни (например, для английского глагола *to be* в Оксфордском словаре) значений соответствующих слов, и формальное регламентирование употребляемости этих слов принципиально невозможно.

¹ См., например, W. Weaver, Translation, «Machine translation of languages», стр. 23.

² См. A. D. Booth, W. N. Locke, Historical introduction, там же, стр. 14.

³ См. Л. И. Жирков, Границы применимости машинного перевода, ВЯ, 1956, № 5.

⁴ М. В. Келдыш, А. А. Лянунов, М. Р. Шура-Бура, указ. статья, стр. 31.

Предлагаемый рядом исследователей¹ способ определения значения слова по непосредственно предшествующему и непосредственно следующему за ним слову вызывает сомнения. В этом случае предполагается, что в программе, вложенной в машину, будут предусмотрены все возможные комбинации данного слова в данном значении с другими словами, а это мало вероятно (кроме как в случае фразеологически связанных значений).

Однако, если даже предположить, что при машинном переводе удастся достигнуть некоторого сужения круга искомых значений слов путем определения частоты употребляемости слова по статистическим данным или путем установления достаточных формальных признаков «микроконтекста»², сфера применения этого метода может быть весьма ограниченной. Во-первых, потому что слов, связанных контекстуально и допускающих только один определенный перевод, можно найти немного, и, во-вторых, потому, что нельзя установить требуемые для каждого перевода границы контекста.

Не поддаются никакому регламентированию и переводы слов, употребленных в переносном значении³.

Сравнительно большое распространение имеют в языке случаи окказионального употребления слова. Например, в английском языке можно весьма часто наблюдать образование прилагательных из других частей речи и целых групп слов путем конверсии; такие прилагательные в большинстве случаев не входят в постоянный словарный состав языка, но образуются, так сказать, «для одного раза», для употребления в одном конкретном контексте.

Такого рода образования, которыми изобилует современная английская публицистическая и художественная литература, а также многочисленные газетные клише типа *Big Four summit meeting* «встреча глав правительств четырех держав», невозможно переводить без знания контекста, а иногда и без знания исторической обстановки, политических взглядов автора и т. д. При этом главная трудность для перевода создается здесь тем обстоятельством, что смысловые связи внутри такого типа сочетаний могут быть весьма сложными и многообразными при одной и той же грамматической структуре словосочетания. Именно здесь содержится многочисленные «подводные камни», преодолеть которые может только «мыслящий», а не «механический» переводчик. Не преодолимую до конца трудность представляет также перевод однотипных грамматических моделей — словосочетаний.

В английском языке совершенно аналогичные внешние сочетания переводятся каждое по-разному. Перевод такого рода сочетаний требует от переводчика тщательного анализа текста с учетом всех его грамматических и смысловых связей, всех возможных вариантов перевода, а также знание всего контекста в целом.

Большое место в работе переводчика занимает и перевод так называемых «реалий», значение которых может быть раскрыто часто только при помощи специальных справочников, энциклопедий и пр. Само собой разумеется, что раскрытие содержания подобных выражений не может быть регламентировано языковыми правилами и не может быть, следовательно, обработано как программное указание для машинной операции.

Серьезная трудность возникает при переводе различного типа фразеологических сочетаний и идиом. Обычно в данном рода случаях предлагают все сочетание вносить в словарь целиком как неделимую единицу (что и имеет место в ныне существующих словарях обычного типа). Однако здесь возникает непреодолимое препятствие — омонимия фразеологических и свободных сочетаний, например англ. *black dog* «черная собака» (свободное сочетание) и «унылые» (фразеологическая единица)⁴.

Многие фразеологические единицы имеют не одно, а несколько значений; например англ. *red herring* может иметь значения: 1) «копченая селедка», 2) «солдат», 3) «что-либо намеренно отвлекающее внимание, сбивающее со следа». Здесь мы снова возвращаемся к известной проблеме многозначности.

Вопросы выбора из ряда синонимов необходимого слова для машины могут быть решены только в случае наличия формально-однозначных разделяющих признаков. Например, при конверсии частей речи в английском языке необходимое слово выбирается по признаку присутствия грамматического форманта (*snow-snows*) или при наличии прямых признаков контекста (проверка на рядом стоящее слово для определения части речи: если за данным словом идет глагол, то предыдущее — существительное). В случае отсутствия непосредственных критериев для выбора значения слова решение может зависеть только от общего контекста (например, значение фразы *French sup-*

¹ См., например, L. E. Dostert, The Georgetown — I. B. M. Experiment, «Machine translation of languages», стр. 129—131.

² См.: E. Reifler, The mechanical determination of meaning, там же, стр. 149—154; W. Weaver, указ. соч., стр. 21.

³ См. об этом подробнее в статье Г. В. Колшанского «К вопросу о возможностях машинного перевода» («Бюллетень Объединения по проблемам машинного перевода», № 4, М., 1957, стр. 82).

⁴ См. А. В. Кунин, Англо-русский фразеологический словарь, М., 1955, стр. 298.

ply troublesin Syria может быть: а) «Французы проводят беспорядки в Сирии»; б) «Затруднения французов в отношении снабжения в Сирии»). Следует также учитывать, что многие авторы в целях достижения большей выразительности иногда видоизменяют структуру фразеологического сочетания, нарушают порядок следования компонентов, вводят в него новые слова и т. д. Проблему сохранения стилистических особенностей при переводе текстов различного жанра мы здесь разбирать не будем.

Итак, природа языка во всех его элементах (лексика, морфология, синтаксис), служащая средством материализации общественного сознания человека, средством общения человека, является по своей внутренней структуре весьма сложным образованием, свойственным только человеку в процессе его общественно-сознательной деятельности. Поэтому механизм языка не может быть доступен целиком методу чисто формального анализа. Законы, управляющие деятельностью мышления и языка, не могут быть сведены к сумме элементарных законов, определяющих формы бытия и движения в неорганическом мире. При всей аналогии функционирования машины с некоторыми мыслительными процессами, протекающими в мозгу, механическая модель никак не может воспроизвести творческую деятельность органа высокоорганизованной живой материи — мозга¹.

Следовательно, все, что выходит за рамки элементарных, формально и однозначно определяемых правил функционирования языка, не может быть объединено схемой и, следовательно, не может быть объектом машинной обработки. Не развитие техники ставит предел возможностям машинного перевода, а сам язык. Машина может сделать лишь то, что ей укажет человек, а в данном случае человек бессильен воспроизвести всю программу возможного функционирования языка в соответствующий момент перевода, ибо для этого ему пришлось бы возвращаться каждый раз к творческой деятельности мозга, другими словами, прибегать к помощи своего мышления при нахождении нужного результата «по смыслу».

Лингвистика не в состоянии изобразить свой предмет — язык — в виде практически конечного числа аксиом, что принципиально доступно только науке, имеющей дело с постоянными величинами, такой, как математика. В данном случае язык представляет собою такую совокупность, которая не поддается во всем объеме алгоритмическому разрешению. Здесь вообще лежит предел для любого вида кибернетических устройств.

Таким образом, несмотря на быстрое и эффективное развитие техники вычислительных машин, можно придать категорическую форму следующему высказыванию самых горячих защитников машинного перевода: «Возможно, что никакая управляющая машина никогда не сможет моделировать мозг, а будет моделировать только выполнение мозгом ограниченного класса функций»².

Нельзя в то же время отрицать практическую ценность машинного перевода; внедрение его в широких масштабах будет полезным мероприятием. Однако необходимо при этом всегда учитывать, что этот вид перевода будет выполнять лишь элементарную функцию, будет играть лишь подсобную роль при переводе к тому же только специальных текстов³. При необходимости получения полноценного перевода машинный перевод может быть лишь исходным материалом для переводчика-редактора, а в необработанном виде или в форме созданного для этой цели «промежуточного» языка может выполнять роль быстро действующего механического информатора.

Задачей лингвистической науки, кооперирующейся в настоящее время с математическим направлением в целях разработки техники машинного перевода, и является сейчас изучение круга элементарных языковых явлений, поддающихся формально-структурному описанию, поддающихся математической обработке. Это направление не может претендовать на роль универсальной методики лингвистического исследования; наоборот, оно само в значительной степени строится на результатах структурного, описательного, семантического и исторического анализа языка, но соприкосновение лингвистики с математическими методами и аппаратом может быть благотворным для нее и приблизит науку о языке к сфере точных наук.

Л. С. Бархударов и Г. В. Колманский

¹ См.: П. К. Анохин, Физиология и кибернетика, ВФ, 1957, № 4, стр. 153—154; И. И. Гальперин, О рефлекторной природе управляющих машин, там же, стр. 165.

² М. В. Келдыш, А. А. Ляпунов, М. Р. Шур-Бура, указ. статья, стр. 37.

³ Л. И. Жирков, указ. статья, стр. 122.