

# ОПЫТЫ МАШИННОГО ПЕРЕВОДА

## НЕКОТОРЫЕ ЛИНГВИСТИЧЕСКИЕ ВОПРОСЫ МАШИННОГО ПЕРЕВОДА

Создание электронных вычислительных машин, автоматически производящих большие серии арифметических и логических операций, поставило перед учеными новую проблему использования этих машин для осуществления перевода текстов с одного языка на другой.

К осуществлению идеи перевода посредством электронной машины приступили ученые А. Д. Бут и У. Уивер в 1946 г. Первоначально они представляли себе перевод лишь как простой подстрочник к тексту; машине в процессе перевода отводилась вспомогательная роль «запоминания» словарных единиц и их выдачи. В 1948 г. Р. Г. Риченс предложил несколько расширить сферу действия машины с тем, чтобы она путем сравнения слова в тексте с соответствующей основой в словаре не только выдавала основу слова на языке перевода, но и по «остатку» (окончанию) могла снабдить основу слова в переводе некоторой грамматической информацией. Начиная с 1948 г. проблемой механического перевода занимаются специалисты по кигайскому языку Э. Рейфлер, специалисты по германским языкам В. А. Освальд и С. Л. Флетчер, специалист по испанскому языку У. Булл, профессор философии И. Бар-Хиллел и другие.

В 1950 г. Э. Рейфлер предложил осуществить перевод по принципу «машина плюс два человека». Согласно этому принципу, при машине предполагалось иметь двух редакторов, знающих только по одному языку: первый редактор («пре-редактор») должен преобразовывать оригинальный текст в удобную для машины форму (т. е. устранять лексическую и грамматическую омонимию, полисемию и т. д.), а второй («пост-редактор») — редактировать полученный при помощи машины перевод (т. е. менять там, где это необходимо, порядок слов, подбирать нужные синонимы и исправлять возможные ошибки)<sup>1</sup>. От идеи «пре-редактирования» сторонники этого принципа впоследствии отказались.

В 1952 г. С. Додд предложил проект «идеального языка», на котором должен быть написан материал, подлежащий переводу. В этом «идеальном языке» порядок слов должен быть строго фиксирован, а окончания — заменены служебными словами<sup>2</sup>. С. Додд полагал, что если к этим двум правилам добавить сравнительно ограниченный словарь и по возможности придерживаться фонетического написания, то проблема механического перевода с такого языка не будет представлять особых трудностей.

Среди специалистов, занимающихся вопросами машинного перевода, существует и другое направление, сторонники которого пытаются решить возникающие при машинном переводе лингвистические проблемы. Так, например, В. А. Освальд и С. Л. Флетчер подробно исследовали вопрос о возможности машинного перевода с немецкого языка на английский с точки зрения синтаксиса этих языков<sup>3</sup>. Аналогичные вопросы перевода с русского языка на английский анализируются в статье К. Харпера, специалиста по славянским языкам<sup>4</sup>. Ряд исследователей (В. А. Освальд и Р. Г. Лоусон, А. Г. Эттингер, И. Бар-Хиллел) рассматривают вопросы, связанные с задачами составления машинного словаря<sup>5</sup>; некоторые ученые (Л. и А. Вундхейлер,

<sup>1</sup> См. E. Reifler, *Studies in mechanical translation*, 1950 (mimeограф. изд.).

<sup>2</sup> S. C. Dodd, *Model English*, сб. «Machine translation of languages», New York — London, 1955.

<sup>3</sup> V. A. Oswald and St. L. Fletcher, *Proposals for mechanical resolution of German syntax patterns*, «The modern language forum», vol. XXXVI, № 3—4, Los Angeles, California, 1951.

<sup>4</sup> K. E. Harper, *A preliminary study of Russian*, сб. «Machine translation of languages».

<sup>5</sup> См.: V. A. Oswald, R. H. Lawson, *An idioglossary for mechanical translation*, «The modern language forum», vol. XXXVIII, № 3—4, 1953, Los-Angeles, California; G. Oethinger, *The design of an automatic Russian-English technical dictionary*, сб. «Machine translation of languages»; еро же, *Distribution of word length in technical Russian*, «Mechanical translation», vol. I, № 3, Cambridge, Mass., 1954; Y. Bar-Hillel, *Idioms*, сб. «Machine translation of languages».

В. Ингве и др.) предпринимают попытки применения логических методов для построения «машинного» синтаксиса<sup>1</sup>.

В начале 1954 г. в США Л. Достертом, П. Гарвином и П. Шериданом впервые был практически осуществлен механический перевод с русского языка на английский на машине ИБМ-10<sup>2</sup>. Таким образом, реальность машинного перевода не подлежит сомнению. В настоящее время вопрос состоит в том, чтобы найти наиболее эффективный путь его осуществления.

Задача настоящей статьи — осветить языковедческие вопросы, которые возникают в связи с разработкой проблемы машинного перевода. Эти вопросы имеют тем более важное значение, что в настоящее время именно они являются основными для решения проблемы машинного перевода<sup>3</sup>.

Для машинного перевода требуется решение двух комплексов языковедческих вопросов, связанных с составлением, во-первых, словаря, а во-вторых, свода грамматических правил. Оба круга вопросов, разумеется, взаимосвязаны. Однако поскольку каждый из них имеет свою специфику, целесообразно рассматривать их отдельно.

Машинный перевод предполагает создание двуязычного словаря особого типа, отличающегося от обычных двуязычных словарей (как общих, так и терминологических) своим словариком и структурой, а также грамматическим оформлением и принципом перевода слова.

Поскольку при помощи машины предполагается сначала переводить только научно-техническую литературу, то на первых порах целесообразно ограничиться составлением отдельных отраслевых словарей. Каждый такой отраслевой словарь должен содержать наряду с терминологическими также и нетерминологические слова, употребляемые в научно-технической литературе данной специальности.

В этой связи представляют интерес исследования В. А. Освальда и Р. Х. Лоусона, составивших словарь по нейрохирургии с применением статистических методов подсчета и отбора слов<sup>4</sup>. Освальд и Лоусон подвергли обследованию 16 специальных статей на немецком языке. Выписанные из первой статьи и распределенные по частям речи слова сопоставлялись поочередно со словами остальных статей. Дополнения вносились до тех пор, пока не были зафиксированы все более или менее употребительные в нейрохирургии слова. Оказалось, что уже примерно с четырнадцатой статьи дополнения вносили очень мало нового.

Обследование частоты употребления слов в специальных текстах является крайне важным для составления машинного словаря. Оно дает возможность правильно отобрать лексический материал. При этом допускается, что некоторое количество слов не будет занесено в словарь и, следовательно, в тексте останется без перевода. Тем не менее не следует перегружать словарь пассивным материалом и этим увеличивать время поиска слов<sup>5</sup>.

Возникает вопрос, в каком виде следует поместить в словарь отобранный материал. Во всех экспериментальных переводах словарь состоял либо полностью, либо частично из основ, а не из грамматически оформленных слов. Правила же словообразования и словосложения, по возможности, следует программировать. Это не всегда удастся, поскольку многие малопродуктивные и непродуктивные словообразовательные средства (например, немецкие суффиксы *-nis* и *-sal*, английский суффикс *-age*) не имеют стандартного перевода в русском языке, т.е. при переводе соответствующих слов употребляются русские слова, оформленные различными суффиксами (ср. нем. *Hindernis* «препятствие», *Finsternis* «темнота»; англ. *drainage* «осушение», *advantage* «преимущество» и т. п.). Поэтому слова с подобными суффиксами правильнее внести в словарь в виде отдельных статей.

<sup>1</sup> L. and A. Wundheiler, Some logical concepts for syntax, сб. «Machine translation of languages»; V. H. Yngve, Syntax and the problem of multiple meaning, там же.

<sup>2</sup> Результаты, полученные при этом, освещались на страницах нашей печати. См.: Реферативный журнал «Математика», 1954, № 10; журналы «Природа», 1955, № 8 и ВЯ, 1955, № 6.

<sup>3</sup> Интересно, что это обстоятельство отмечается и руководителем первого опыта машинного перевода в США Л. Достертом (см. L. Dostert, The Georgetown — I. B. M. experiment, сб. «Machine translation of languages», стр. 124).

<sup>4</sup> См. V. A. Oswald, R. H. Lawson, указ. соч.

<sup>5</sup> При составлении словаря для машинного перевода можно пользоваться уже готовыми подсчетами частотности слов и включать в словарь нетерминологическую лексику с наибольшими показателями частотности. Интересно, например, «Подсчет ходовых слов русского литературного языка» под ред. Г. Джоссельсона (Н. Н. Josselson, The Russian word count, Detroit, 1953), где указывается частотность 2 тысяч слов, встречающихся в произведениях художественной литературы, в публицистике и научной литературе.

Вопрос о стандартизации перевода не менее важен и для наиболее продуктивных суффиксов, так как одному иностранному суффиксу почти всегда соответствуют несколько русских. Так, нем. (или англ.) суффиксы *-er* может соответствовать любой из русских суффиксов, образующих *nomina agentis*: *-ец* (и его производные), *-ик* (*-ник*), *-щик* (и его производные), *-тель* (*-итель*), *-ач* и т. п. Даже в тех случаях, когда какому-либо иностранному суффиксу чаще всего соответствует определенный суффикс русского языка (ср. например, нем. *-ung* и русск. *-ение*), все же имеется очень много исключений, которые приходится учитывать. В настоящее время можно наметить следующие пути установления соответствий в области словообразования: во-первых, основы нестандартным переводом словообразующего суффикса (как непродуктивного, так и продуктивного) можно вводить в словарь целиком; во-вторых, переводящие основы в словаре можно снабдить специальными пометами, указывающими, при помощи какого суффикса от данной основы образуется то или иное новое слово.

Особой задачей (по крайней мере, для машинного перевода с немецкого языка) является составление правил перевода сложных слов<sup>1</sup>. При отсутствии таких правил словарь не дал бы возможности переводить сложные слова типа *Gefrierpunktserniedrigung* «понижение точки замерзания», *Bleidioxydschicht* «слой перекиси свинца» и т. п., которые очень распространены в научно-технической литературе. Между тем перевод этих сложных слов производится по элементарнейшим правилам, допускающим формализацию<sup>2</sup>. Важность разработки правил словообразования состоит в том, что при их помощи машина в ряде случаев сможет дать перевод неологизмов.

Одним из наиболее сложных вопросов, связанных с составлением словаря, является вопрос однозначности перевода. Для достижения максимальной однозначности в словаре не должны учитываться различные стилистические синонимы того или иного переводящего слова, а иногда и идеографические синонимы (если они не имеют особого значения для данной области науки и техники). В случае, когда одно и то же слово в разных областях научной литературы имеет различный перевод, переводное соответствие можно снабдить специальной пометой и при введении текста в машину вводить соответствующую информацию о специфике текста. Например, англ. *propagation* обычно переводится как «распространение», а с пометой (биол.) — как «размножение».

Одним из средств достижения однозначности является введение в словарь сочетаний, в которых данное слово имеет перевод, сильно отличающийся от основного переводящего соответствия. При выборе соответствий необходимо учитывать не только частоту их употребления, но и стандартность образования их грамматических форм. Например, для немецкого глагола *aufnehmen*, помимо переводящего слова, выражающего его основное значение, и перевода некоторых сочетаний с этим глаголом (например, *den Betrieb aufnehmen* «начать работу»), нужно дать и другие его переводные соответствия: «вбирать (в себя)», «поглощать». При этом первое соответствие придется, по-видимому, исключить, так как от глагола «вбирать» трудно образовать причастие страдательного залога, а это в ряде случаев будет необходимо (ср. *Nunmehr wird die ätherische Lösung von aufgenommen Wasser befreit* — из учебника химии). Рациональнее принять второе соответствие, так как от глагола «поглощать» указанное причастие может быть образовано.

Наконец, могут встретиться случаи, когда общелитературное слово и омонимичный ему термин одновременно употребляются в данной области научно-технической литературы [ср., например, нем. *Lösung* «решение» и (хим.) «раствор»]. Для правильного выбора переводящего слова нужно дать его в сочетании с другим словом из текста (*Lösung der Aufgabe, der Frage*); если же это окажется трудно выполнимым, придется указать оба варианта перевода и предоставить читателю выбор нужного значения. Таким же образом, видимо, придется поступить и с рядом многозначных терминов типа франц. *courant* «ток, течение, поток».

Как явствует из вышеизложенного, вопрос о переводе слова связан с проблемой сочетаемости. Перевод фразеологически связанных словосочетаний представляет значительные трудности, так как подобные словосочетания не имеют дословных соответствий в различных языках. Возможно несколько способов перевода так называемых идиоматических словосочетаний:

1. Составить отдельный словарь, устанавливающий соответствие между идиоматическими словосочетаниями двух языков. Тогда машина должна начинать перевод

<sup>1</sup> Значительный интерес представляет в этой связи статья Э. Рейфлера (E. Reifler, *Mechanical resolution of the constituents of German substantive compounds*, «Mechanical translation», vol. II, № 1, 1955).

<sup>2</sup> Необходимо предусмотреть случаи, в которых применение подобных правил может привести к абсурду (ср., например, *Stickstoff* «азот», а не «удушливый газ» и т. п.). Во избежание этого соответствующие основы в словаре следует снабдить пометами, указывающими, с какими соединительными элементами они выступают в словосложении, и тем самым предупредить возможность неверного перевода слов типа *Arbeitsamt* «биржа труда» (*Arbeit + s + Amt*), а не «рабочий бархат» (*Arbeit + Samt*).

проверкой каждого словосочетания в этом словаре. Если в тексте не найдено ни одного из зафиксированных в словаре словосочетаний, то перевод осуществляется пословно.

2. Снабдить слово (например, франц. *difference* «разница») определенным знаком, указывающим на сочетаемость слова с другими словами; посредством этого знака отослать машину к следующей статье, в которой приводится данное слово в определенном словосочетании (например, *difference de phase*) и русский эквивалент этого словосочетания («разность фазы»).

3. Использовать предложенный И. Бар-Хиллелем метод исключительно пословного перевода идиоматических словосочетаний<sup>1</sup>. Бар-Хиллел полагает, что в подавляющем большинстве случаев пословный перевод будет удовлетворительным, особенно если учитывать, что определенные словосочетания имеют соответствующие кальки в нескольких языках, например франц. *perdre la tête* и русск. *потерять голову*. Подобные словосочетания могут быть разложены на компоненты без указания на сочетаемость. Однако способ пословного перевода не разрешает всех трудностей, возникающих при переводе идиоматических словосочетаний.

В словаре необходимо учесть случаи омонимии фразеологически связанных и свободных словосочетаний. Например, нем. *Allen Mann erschlagen* в качестве идиоматического словосочетания, употребляющегося в горнозаводском деле, на русский язык переводится как «попасть на старую выработку». Между тем возможен и пословный перевод этого словосочетания — «убить старого человека». Очевидно, некоторые идиоматические словосочетания должны сопровождаться в словаре указаниями на возможность также пословного их перевода.

При анализе машинной переводимого текста возникает необходимость определения грамматической формы, в которой слово выступает в данном контексте. Машина должна располагать сведениями о том, какие формы возможны у данного слова. С этой целью все основы слов, вводимые в словарь, следует снабдить грамматическими пометами и в первую очередь — сведениями о том, к какому разряду слов принадлежит данное. Внутри разрядов слов выделяются группы в зависимости от типов образования тех или иных грамматических форм. Например, внутри глагола различаются группы, по-разному образующие временные формы; внутри существительного — группы, по-разному образующие множественное число, и т. д.

При переводе с французского языка необходимо знать о каждом глаголе, к какой группе (по типу образования форм времени и наклонения) он принадлежит. Традиционно-грамматическое деление глаголов на три группы при этом оказывается недостаточным, так как в третью группу объединяются фактически очень различные типы. Так называемую третью группу глаголов целесообразно разбить на несколько групп, внутри которых формы времени и наклонения образуются единообразно. Выделение новых групп уместно осуществлять в том случае, если формы с чередованием звуков в основах и супплетивностью считать различными вариантами одной и той же основы и давать их в словаре отдельно. Тогда многие глаголы будут представлены несколькими основами (например, глагол *écrire* — двумя: *écri*, являющейся основой инфинитива, причастия прошедшего времени, ед. числа настоящего изъявительного, будущего времени, и *écriv*, являющейся основой причастия настоящего времени, мн. числа настоящего изъявительного, прошедшего простого времени, настоящего сослагательного). В то же время выясняется невозможность отнести некоторые глаголы (из-за специфики образования отдельных их форм) к какой-нибудь из выделенных групп. Специфические формы таких глаголов представляют собой исключения и, очевидно, должны содержаться в словаре в виде отдельных словоформ. Согласно этому, глагол *faire* в словаре должен быть представлен группой основ: *fai, fais, fe, fass* и отдельными словоформами *faites* — 2-е лицо мн. числа наст. времени изъяв. наклонения и *font* — 3-е лицо мн. числа наст. времени изъяв. наклонения.

После того как для слова в тексте отыскана соответствующая основа в словаре, грамматические пометы при этой основе отсылают к таблицам, в которых перечислены окончания, возможные у данного слова. Путем сравнения окончания, оставшегося после отсечения основы анализируемой словоформы, с окончаниями, содержащимися в таблице, устанавливается, какое грамматическое значение передает данная словоформа. Например, в переводимом тексте встретилась форма *écrivons*. В словаре содержится две основы: *écri* и *écriv*. После сравнения с данными словаря от *écrivons* отсекается большая по длине основа, в результате чего остается окончание *-ons*. В словаре содержатся также указания на то, к какой группе относится этот глагол. В отдельной таблице отмечены все окончания глаголов этой группы. Среди них путем сравнения отыскивается окончание *-ons*. В таблице указано, что оно соответствует 1-му лицу множественного числа настоящего времени изъявительного наклонения. Таким образом, мы получили полную информацию о данной словоформе.

Следует оговорить один очень существенный момент. В таблицу грамматических

<sup>1</sup> Y. Bar-Hillel, Some linguistic problems connected with machine translation, «Philosophy of science», vol. 20, № 3, Baltimore-2, Md., 1953.

форм включаются не все формы, возможные для данного разряда слов, но лишь достаточно часто встречающиеся. Поэтому прежде чем решить вопрос о том, какие грамматические формы следует учесть в грамматических правилах, нужно провести анализ частоты их употребления<sup>1</sup>. Малоупотребительные грамматические формы можно не включать в программу для машины. Так, при переводе с французского, по-видимому, можно не вносить в грамматические таблицы такую глагольную временную форму, как *imparfait du subjonctif*, встречающуюся в языке научной литературы очень редко, за исключением только 3-го лица единственного числа.

В отношении грамматики задача перевода состоит в том, чтобы значения, выражаемые морфологическими и синтаксическими категориями языка оригинала, возможно точнее передать грамматическими категориями того языка, на который осуществляется перевод. Эта задача возникает и в условиях обычного перевода, но при машинном переводе пути решения ее представляют ряд особенностей, которые определяются тем, что машина не может оперировать смысловыми связями, имеющимися в тексте<sup>2</sup>. В силу этого фраза при машинном переводе должна быть подвергнута анализу только по формальным признакам. Говоря о формальном анализе явлений языка, мы должны помнить, что именно такой анализ в принципе всегда имеет место при разборе текста на иностранном языке человеком, который плохо владеет этим языком<sup>3</sup>. Весь процесс перевода складывается из двух моментов: из анализа текста по структурным признакам и из построения перевода на основе информации, полученной при анализе.

Возможности формального лингвистического анализа существенно различаются в зависимости от структуры языка. Соответственно различным будет и подход к проблеме машинного перевода с разных языков. В языках синтетических большую долю информации о том, в какой синтаксической функции и грамматической форме выступает то или иное слово, несут словозначительные аффиксы. Русский язык, например, располагает богатыми формальными возможностями, поэтому в нем очень легко строить однозначные в формальном отношении предложения<sup>4</sup>. В немецком языке формальных возможностей меньше, так как имеет место значительная омонимичность формантов, например *-(e)n* является окончанием всех косвенных падежей существительных слабого склонения (*des Menschen*), суффиксом множественного числа существительного (*die Frauen*), окончанием прилагательного (*die guten Frauen*), окончанием 1-го и 3-го лица множественного числа глагола (*marschieren*), формами инфинитива и т. п. Формальный анализ в английском языке представляется еще более сложным вследствие омонимичности многих аффиксов (например, *-(e)s, -ing, -ed*) и явления конверсии. Так, например, словоформа *moves* может обозначать либо множественное число существительного («движения — шаги»), либо 3-е лицо единственного числа настоящего времени глагола («движется»). Поэтому для такого сочетания, как *Soviet peace moves*, однозначный формальный анализ невозможен.

Из вышеизложенного следует, что в области грамматики одной из основных лингвистических проблем, возникающих при машинном переводе с западноевропейских языков на русский, является проблема грамматической омонимии.

<sup>1</sup> Ср. данные о частотности грамматических категорий и форм русского языка, приведенные Г. Джоссельсоном (см. Н. Н. J o s s e l s o n, указ. соч.). Автор приводит показатели частоты употребления отдельных падежных форм существительного, степеней сравнения прилагательного, глагольных видов (соверш. вид. 57,5%, несоверш. — 42,5%), форм наклонения (изъявительное соверш. вида 85,4%, повелительное соверш. вида 0,4%, сослагательное соверш. вида 1%, неопределенная форма соверш. вида 13,2%; изъявительное несоверш. вида 85,7%, повелительное несоверш. вида 0,5%, сослагательное несоверш. вида 0,3%, неопределенная форма несоверш. вида 13,5%), временных форм (наст. вр. несоверш. вида 33,6%, прош. вр. несоверш. вида 66,1%, будущ. вр. 0,3%), форм лица и пр.

<sup>2</sup> Вопросу о влиянии контекста на однозначность перевода посвящена статья А. Каплана (A. K a p l a n, An experimental study of ambiguity and context, «Mechanical translation», vol. II, № 2, 1955).

<sup>3</sup> См. Л. В. Ш е р б а, Перспективы преподавания иностранных языков в школе, сб. «Иностр. язык в школе», М., 1944, стр. 9.

<sup>4</sup> Широкие формальные возможности русского языка способствовали успеху первого опыта машинного перевода (с русского языка на английский). Наличие развитой системы падежных окончаний позволило установить такие соответствия, как *-in* например, в форме *отношением*) — *by* перед знаменательным словом (*by relation*), *-y* (например, в форме *радиусу*) — *to* и т. п. Подобные правила, однако, имеют ограниченную приложимость. Они неприменимы для перевода с английского языка на русский. При более детальной разработке практики перевода с русского языка возникает проблема многозначности формантов. В частности, вопрос о многозначности-ся в глаголе (около 20 значений!) уже привлек внимание исследователей. См. J. W. P e r r y, A practical development problem, сб. «Machine translation of languages».

Для устранения грамматической омонимии необходимо разбить все слова на определенные разряды, или классы, целиком исходя при этом из формальных признаков. При чисто формальном распределении слов по классам слова, основы которых могут образовывать без аффиксации различные части речи (например, нем. *gut* — прилагательное и наречие, англ. *turn* — глагол и существительное), и слова, основы которых могут образовывать только одну часть речи (нем. *eisern* «железный», англ. *establish* «устанавливать»), должны быть отнесены к разным классам. В словаре основы с многозначной морфологической характеристикой должны нести соответствующую помету: например *gut* — «прилагательное-наречие», *turn* — «существительное-глагол» и т. п. Это дает возможность сформулировать правила, которые позволяют по окружению определять принадлежность соответствующего слова к той или иной части речи и лишь после этого осуществлять перевод.

Задача формального анализа требует учета еще одного важного фактора, каким является порядок слов. Проблема порядка слов возникает в случае омонимии членов предложения, например формального совпадения подлежащего и прямого дополнения в английском, немецком и французском языках. Сложными и до сих пор не получившими решения являются вопросы о том, как далеко могут отстоять друг от друга синтаксически связанные слова, а также вопрос об инверсии, наличие которой может привести к полному искажению смысла переводимого текста.

Для формального анализа фразы можно использовать ряд вспомогательных приемов, в том числе прием снятия многозначности, предложенный И. Бар-Хиллелем<sup>1</sup>. Суть этого приема состоит в следующем. Бар-Хиллел выделяет два опорных класса слов: имена и предикативные слова, а для всех остальных слов указывает возможность вхождения их в атрибутивные или предикативные сочетания. Слова, относящиеся к классу имен, обозначаются знаком «и», слова, относящиеся к классу предикативных слов, — знаком «s». Для того чтобы указать, в какое словосочетание (атрибутивное или предикативное) входит данное слово, этому слову в качестве своеобразного «знаменателя» присписывается соответственно знак «и» или «s», который Бар-Хиллел называет аргументом данного элемента. Если некоторый элемент (т. е. слово) связан с другим элементом, стоящим от него справа, то аргумент первого ставится в квадратных скобках; если же он связан с элементом, стоящим от него слева, то его аргумент заключается в круглые скобки. Так, предложение *Poor John sleeps* изображается формулой  $\frac{n}{[n]} \frac{s}{(n)}$ . После составления формулы Бар-Хиллел предлагает, отвлекаясь от конкретных значений элементов, поступить с полученной цепочкой символов, как с дробями, т. е. сокращать их. Аргументы, стоящие в квадратных скобках, сокращают вправо, а аргументы, стоящие в круглых скобках, — влево. После сокращения должно остаться одно «s» (предикативная связь). Иной результат будет указывать на отсутствие синтаксической связанности элементов данной последовательности.

Методом подобной индексации в принципе можно выразить какие угодно сложные связи. Например, такая последовательность, как  $\frac{s}{(n)} \frac{n}{[n]} \frac{s}{(n)} \frac{n}{(n)} \frac{s}{[n]} \frac{n}{[n]} \frac{n}{[n]}$ , при постепенном упрощении дает «s». Действительно, данная последовательность может символизировать предложение типа: *John knew that Paul was a poor man*.

Наиболее ценным в приеме Бар-Хиллела является указание на то, что целый ряд одних и тех же слов в разных окружениях может иметь разные индексы, т. е. входить в словосочетания различного грамматического типа. Поочередно присписывая словам фразы различные индексы, можно установить принадлежность этих слов к грамматическим классам и тем самым сократить степень неоднозначности перевода, а в некоторых случаях достичь однозначности.

Одним из важных общих вопросов машинного перевода является вопрос о выделении отрезков текста, в пределах которых действуют составленные для машины правила. В качестве такого отрезка удобнее всего взять предложение, границы которого отмечаются знаками препинания: точкой, точкой с запятой, восклицательным знаком, вопросительным знаком. Однако вопрос усложняется тем, что употребление знаков препинания не всегда бывает однозначным. Так, точка ставится после порядковых числительных в немецком и английском языках; восклицательный знак используется в математике как знак факториала (ср.  $5! = 1 \times 2 \times 3 \times 4 \times 5$  и т. п.). Кроме того, между двумя точками может стоять очень длинный период, перевод которого из-за большого количества слов разного типа может представлять значительные трудности. Поэтому возникла необходимость дать формальные правила употребления знаков препинания, включая запятую. Попытка выработки подобных правил была предпринята в отношении немецкого языка. Были составлены правила, устанавливающие, в каком окружении запятая является концом предложения (boundary mark), и позволяющие

<sup>1</sup> Y. Bar-Hillel, A quasi-arithmetical notation for syntactic description, «Language», vol. 29, № 1, Baltimore-2, Md., 1953.

ограничить отрезок текста, непосредственно подлежащий переводу<sup>1</sup>. Кроме знаков препинания необходимо было найти формальные структурные признаки законченности отрывка. Такими признаками, по-видимому, могут служить рамочная конструкция (только для немецкого языка), подчинительные или сочинительные союзы в последующем тексте и т. п.

В заключение необходимо остановиться на вопросе о границах машинного перевода. Поскольку не все вопросы словаря и грамматики можно решить чисто формально, в переводе неизбежно будет оставаться непереуведенным незначительный процент слов и выражений, а также будут иметь место некоторые погрешности против нормы сочетаемости слов и стилистического употребления грамматических конструкций. Не приходится говорить о том, что проблемы, которые по самой своей природе требуют творческого подхода, также не могут быть решены в процессе машинного перевода. Признание этих ограничений отнюдь не ставит под сомнение самую идею машинного перевода, а только еще раз подтверждает, что машины могут выполнять задачи такого рода, облегчая труд человека, но не заменяя его.

*Т. Н. Молошная, В. А. Пурто, И. И. Ревзин, В. Ю. Розенцвейг*

<sup>1</sup> V. A. Oswald and St. L. Fletcher, указ. соч.