

Т. М. ТЕР-МИКАЭЛЯН, Р. Л. УРУТЯН

ОБЩЕЕ ОПИСАНИЕ МАШИНЫ «ГАРНИ» И ПРОЦЕССА РЕАЛИЗАЦИИ АЛГОРИТМА ПЕРЕВОДА

В настоящей статье дается описание устройства и функционирования специализированной электронной цифровой машины «Гарни», предназначенной для автоматизации перевода с одного языка на другой. В первом параграфе дается краткое описание основных свойств алгоритма автоматического перевода. В двух последующих параграфах описываются специфические узлы машины «Гарни», свойства которых вытекают из требований, предъявляемых к специализированной машине характерными особенностями алгоритма перевода. Последний параграф посвящен описанию блок-схемы машины и самого процесса перевода. В настоящее время на машине «Гарни» осуществляется русско-армянский перевод математических текстов, хотя авторам представляется, что она спроектирована в расчете на реализацию перевода для любой пары языков. При этом имеются в виду алгоритмы, опирающиеся на анализ и синтез поверхностной синтаксической структуры переводимой фразы.

§ 1. Структура алгоритма

С точки зрения характера обработки информации весь процесс автоматического перевода может быть разбит на два этапа: морфологический и синтаксический. Каждый из этих этапов, в свою очередь, делится на два: анализ и синтез. Охарактеризуем вкратце эти этапы (блок-схема алгоритма перевода изображена на рис. 1).

Морфологический анализ заключается в сопоставлении каждому слову переводимого текста (на языке оригинала) некоторой грамматической информации, называемой его характеристикой [1]. В характеристику входят словарный номер этого слова (или номера, если его перевод неоднозначен) и его грамматические признаки (типа—часть речи, падеж, время, характер управления и т. п.). Наоборот, морфологический синтез заключается в конструировании слова на языке перевода, исходя из номера этого слова и его грамматических свойств, выработанных грамматическим анализом фразы на языке оригинала.

Информация, вырабатываемая в результате морфологического анализа, может быть неоднозначной (омонимичной). Эта омонимичность

выражается в том, что слово получает не один номер, а столько, сколько переводов учитывает данный алгоритм (два значения слова «следует» во фразах: «следует за» и «из...следует»), а также и в том, что одно слово получает несколько различных признаков (слово «прямая» может быть прилагательным или существительным, или, как говорят, слово является омонимом типа существительное-прилагательное), из которых в

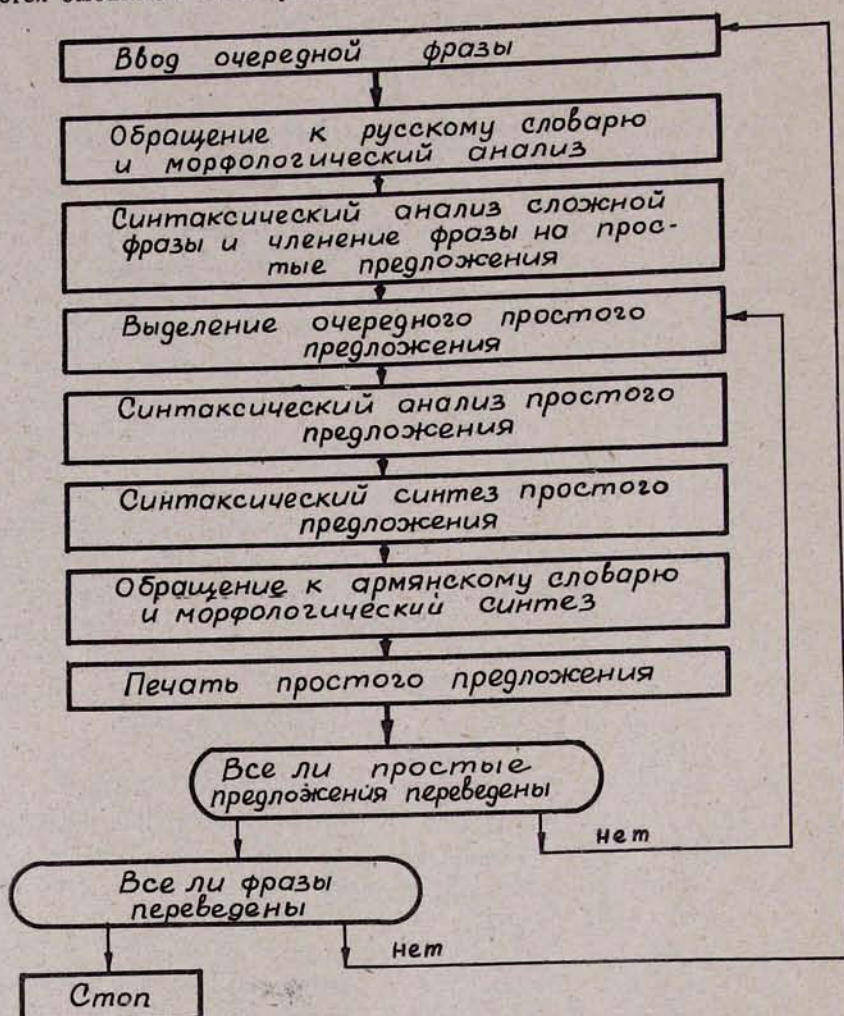


Рис. 1.

данном контексте реализуется только один. Устранить эту омонимию призван синтаксический анализ, в задачу которого входит также выявление связей-управлений между словами фразы (например, существительное управляет определением-прилагательным: «кратный интеграл», «кратная производная»). В задачу синтаксического синтеза входит выражение этих связей средствами языка перевода (выработка

грамматических признаков всех слов, перемена порядка слов, добавление одних—например, при образовании сложных временных форм—и исключение других служебных слов).

Задача морфологического анализа может быть решена двумя различными способами. Один заключается в том, что в запоминающее устройство машины записывается словарь, содержащий все формы слов: например, *интеграл, интеграла, интегралу...*, *вычисляю, вычисляешь, вычисляет...* и т. д. В этом случае форма «интегралу» будет переводиться так: в словаре будет найдена форма, полностью совпадающая с анализируемой, и из информации к этой форме выяснены ее признаки, а также словарный номер (или номера), по которому, при морфологическом синтезе, будет найден армянский эквивалент этого существительного.

Несмотря на свою простоту (существенно упрощается анализ формы слова), этот способ крайне затруднителен тем, что требует запоминающих устройств (ЗУ) большого объема, создание и надежная эксплуатация которых сопряжены в настоящее время с рядом технических трудностей.

При втором способе словарь содержит лишь основы* слов, и при морфологическом анализе отыскивается максимальная по длине основа, целиком вкладывающаяся в данное слово. Далее, от слова отсекается основа, а образовавшийся остаток анализируется в тех местах таблиц суффиксов и окончаний, которые указаны в информации к основе. Из той же информации извлекаются сведения о словарном номере анализируемого слова и вообще любые сведения, которые требуются для реализации алгоритма. Естественно, что результат анализа слова при обоих методах совпадает.

Этот способ позволяет от одной основы получить не только различные грамматические формы одного слова, но и проанализировать различные слова, имеющие один корень. Таков принцип морфологического анализа, на котором мы остановили свой выбор в силу его экономичности. Вместе с методом ассоциативной выборки (см. ниже, § 2) это позволило закодировать русский математический словарь объемом в 6 тысяч слов (3526 основ) с помощью 360 000 бит (о структуре алгоритма морфологического анализа см. подробнее в [1]).

Хотя синтаксический анализ и синтез решают противоположные задачи, и тот и другой весьма близки по характеру выполняемых ими преобразований. Они состоят из большого набора лингвистических правил, каждое из которых в основном проверяет, в какой последовательности идут во фразе слова с теми или иными признаками. В качестве примера мы рассмотрим следующее, несколько упрощенное правило. Пусть во фразе имеются три слова, первое из которых является омонимом типа существительное—прилагательное и типа именительный—ви-

* В отличие от общепринятой терминологии мы обычно в понятие основы не включаем суффикс (см. [1]).

винительный падеж, второе является переходным глаголом и третье—существительным-омонимом типа именительный—винительный падеж. Пусть, далее, между двумя крайними из этих трех слов нет больше других слов, обладающих теми же свойствами, что и эти два слова. Тогда первое из этих слов является подлежащим, то есть существительным в именительном падеже, второе—сказуемым, а третье—прямым дополнением, то есть существительным в винительном падеже. Как видим, в этом правиле нас в первую очередь интересует последовательность, в которой расположены слова во фразе, имеющие указанные свойства. Убедившись, что такая последовательность слов действительно имеется, мы изменим грамматические признаки (вычеркнем одни и впишем другие) у соответствующих слов и, кроме того, запишем оказуемому, что им управляет подлежащее, а дополнению—что им управляет сказуемое. Чтобы проиллюстрировать это правило, рассмотрим табл. 1.

Таблица 1

Порядковый номер слова	Словарный номер слова	Существительное	Прилагательное	Предлог	Знак препинания	Глагол	Переходность	Деепричастие	Формула	Именительный падеж	Винительный падеж	Родительный падеж	Сущ. или прилат. имен. или вин. падеж	Переходный глагол	Сущ. имен. или винит. падеж	Подлежащее	Сказуемое	Прямое дополнение	Номер хозяина
1	N ₁	1	1							1	1		1			1			
2					1														
3	N ₃							1											
4	N ₄			1															
5	N ₅	1								1		1							
6	N ₆	1										1							
7						1													
8	N ₈						1	1						1			1		1
9	N ₉	1								1	1				1			1	8
10	N ₁₀			1															
11	N ₁₁	1									1								
12									1										
13					1														

в которой слова переводимой фразы (включая и знаки препинания) записаны столбцом друг под другом. На этой таблице изображена матрица, столбцы которой соответствуют различным грамматическим признакам, а строки—характеристикам. Характеристики состоят из единиц или нулей (нами для простоты не выписанных), в зависимости от того, обладает или нет данное слово данными признаками. Столбцы полученной матрицы называются шкалами соответствующих признаков. Глядя на эту матрицу, можно сказать, что в данной фразе существительное находится на первом, пятом, шестом, девятом и одиннадцатом местах, глагол находится только на восьмом месте и т. д. Можно строить и более сложные шкалы, например шкалу переходных глаголов, шкалу существительных в именительном или винительном падежах, шкалу су-

ществительных или прилагательных в именительном или винительном падежах и т. д. (см. шкалы 14—16 в табл. 1). Введение шкал существенно упрощает реализацию правил синтаксиса. Например, из анализа шкал 14—16 (табл. 1) видно, что в первой строке таблицы (у слова «прямая») надо погасить признаки «прилагательное» и «винительный падеж» и выработать признак «подлежащее», в восьмой строке выработать признак «сказуемое», а в девятой строке надо погасить признак «именительный падеж» и выработать признак «прямое дополнение». При этом можно даже не знать, что это десятая строка, а достаточно знать, что это та самая строка, в которой в столбце «существительное в именительном падеже» стоит единица. Кроме того, восьмому слову (тому, у которого в столбце сказуемого стоит единица) мы впишем в шкалу «номер управляющего слова» номер той строки (в нашем случае первой), у которой в столбце подлежащего стоит единица, и, аналогично, девятому слову будет вписано, что им управляет восьмое слово.

Важно отметить, что в процессе синтаксического анализа и синтеза мы оперируем в основном со шкалами признаков так, как со столбцами полученной матрицы, хотя и здесь приходится иметь дело с ее строками при выполнении таких операций, как перестановка слов, вписывание новых слов, исключение некоторых слов, вписывание номеров хозяев и т. п.

Точно так же в целях экономии объема ЗУ в морфологическом синтезе мы остановились на том варианте, при котором по словарному номеру синтезируемого слова из армянского словаря извлекается основа, а из записанной к ней информации о правилах словообразования и по выработанным в синтаксисе грамматическим признакам извлекаются требуемые суффикс и окончание (из соответствующих таблиц); так конструируется нужная в данном случае форма слова. Более того, так же, как в русском анализе, при синтезе от одной основы образуются не только различные формы одного и того же слова, но и формы различных слов, имеющих общий корень. Иными словами, от одной основы могут образоваться слова, имеющие различные словарные номера. Поэтому для осуществления поиска в начале статьи ставится наименьший из номеров тех слов, которые образуются от данной основы, а в информацию к этой основе записывается, как образуется каждое слово в отдельности (подробнее о структуре армянской словарной статьи см. ниже, § 2). Армянский математический словарь в кодированном виде занимает 310 000 бит.

Заметим, что в разработанном у нас алгоритме морфологического синтеза широко используются шкалы некоторых признаков.

§ 2. Запоминающие устройства машины

Как мы видели, в процессе машинного перевода приходится оперировать то с характеристиками слов, то со шкалами признаков. В универсальных машинах для хранения одних и других используются различные ячейки ЗУ. Однако изменение информации в одних из них тре-

бует соответствующего изменения в других, что осуществляется с помощью достаточно провоздких программ. Чтобы избежать этого, решено было сконструировать специальное ЗУ, названное двумерным оперативным запоминающим устройством (ДОЗУ) [2].

Принципиально оно представляет собой плоскую прямоугольную матрицу двоичных элементов памяти, содержащую 128 строк и 256 столбцов. (Размеры матрицы были выбраны исходя из реальных нужд алгоритмов МП достаточно широкого класса). Метод объединения этих элементов памяти таков*, что позволяет обращаться к ДОЗУ (читать и записывать) как по строкам, так и по столбцам матрицы. Поскольку регистры операционного устройства машины «Гарни» содержат 32 разряда (см. ниже, § 3), все ДОЗУ разбито на квадратные подматрицы-секции размерностью в 32×32 разряда (см. рис. 2). Эти секции нуме-

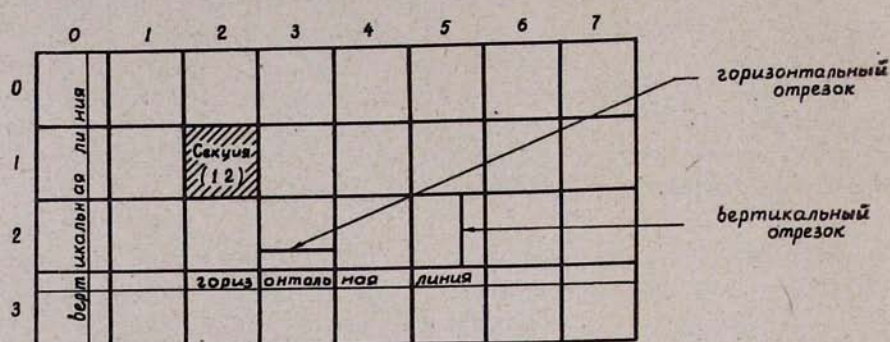


Рис. 2. Двумерное оперативное запоминающее устройство.

руются с помощью двух координат, дающих соответственно порядковые номера секции слева направо и сверху вниз. Тридцатидвухразрядные горизонтальные и вертикальные линии этих секций называются отрезками и нумеруются числами от 0 до 37 (имеется в виду восьмеричная система) в порядке их следования сверху вниз и слева направо. Таким образом, для описания** каждого отрезка достаточно иметь 11 двоичных разрядов (см. рис. 3), из которых правые пять изображают номер отрезка в секции, три последующие изображают горизонтальную, а следующие два—вертикальную координаты секции и, наконец, один разряд служит признаком горизонтального или вертикального направления отрезков. Важно отметить, что при переходе, например, от последнего, тридцать первого горизонтального отрезка некоторой секции к нулевому горизонтальному отрезку секции, расположенной непосредственно ниже первой, меняется вертикальная координата секции. Следовательно, если в младших пяти разрядах адреса стоит число 37, то

* Подробнее об этом см. статью А. М. Амбарцумяна «Двумерное оперативное запоминающее устройство» в настоящем сборнике.

** Подробнее об адресации ячеек ДОЗУ см. в статье А. Г. Ованисяна «Формирование исполнительных адресов запоминающих устройств» в настоящем сборнике.

прибавление к нему 1 в младший правый разряд должно вызвать передачу единицы переноса из левого (пятого) разряда не в соседний с ним слева разряд, а в младший разряд вертикальной координаты секции (см. рис. 3), т. е. должен быть переброшен через три разряда на-лево*.

Перенос единицы при переходе от одной секции расположенной под ней

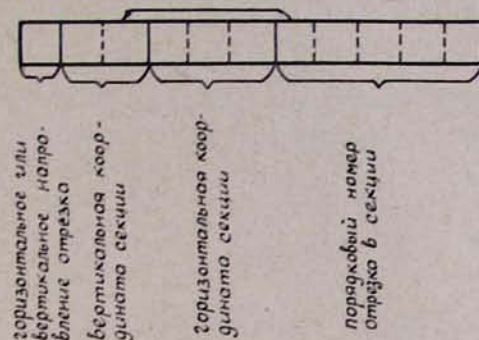


Рис. 3.

Описанное ДОЗУ является единственным оперативным (допускающим программные чтение и запись) ЗУ, имеющимся в машине. В нем хранится обрабатываемая фраза. Вся остальная информация хранится в пассивном запоминающем устройстве (ПЗУ)**. Оно выполнено на кассетах, по 2048 тридцатишестиразрядных ячеек каждая (информация в ПЗУ может быть введена только предварительно). В настоящем варианте машина содержит восемь кассет, из которых три отведены под хранение правил алгоритма русско-армянского перевода, две—под хранение программ и две—всех лингвистических таблиц (таблиц суффиксов, окончаний и т. п.). Одна кассета—резервная.

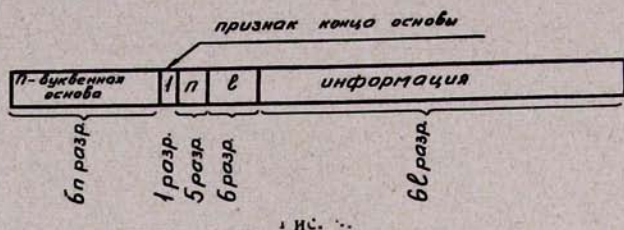
Существенно отметить, что использование ПЗУ особенно выгодно тогда, когда объем информации не слишком велик. Компактная запись правил алгоритма перевода и программ стала возможной только благодаря наличию в машине «Гарни» ДОЗУ и целого ряда специальных команд, выполняемых операционным устройством машины (см. ниже, § 3).

Как указывалось, русский словарь состоит из словарных статей, каждая из которых содержит основу и информацию к ней. Под код каждой буквы выделено шесть разрядов, причем левый разряд равен нулю (коды тридцати двух букв русского алфавита изменяются от 000000 до 011111 в алфавитном порядке букв). Поскольку различные основы содержат, вообще говоря, различное число букв, после основы ставится

* Подробнее об этом см. в статье А. Г. Ованисяна «Формирование исполнительных адресов запоминающих устройств» в настоящем сборнике.

** Подробнее об устройстве ПЗУ см. в приложениях к настоящему сборнику.

специальная секста (так мы называем шестерку двоичных разрядов), в которой левый разряд равен единице, а в остальных пяти—кодируется число, равное числу букв в основе. Следующая секста используется для записи числа секст, входящих в информационную часть словарной статьи, идущей после этой сексты. Общий вид статьи изображен на рис. 4. Все статьи словаря расположены в алфавитном порядке основ, с тем лишь исключением, что из двух основ, вкладывающихся одна в другую, большая предшествует меньшей.



Для хранения словарей в машине «Гарни» используются два магнитных барабана. Поскольку каждая дорожка барабана содержит 8500 двоичных разрядов, весь словарь разбит на части, объем которых при кодировании не превышает этого числа. На каждой дорожке все словарные статьи следуют непосредственно одна за другой (на дорожках нет адресов!). Такое расположение статей дает громадный выигрыш в использовании объема запоминающего устройства, но значительно осложняет поиск нужной статьи по заданному слову. Этот поиск построен по ассоциативному методу, заключающемуся в следующем. По коду первых четырех букв очередного переводимого слова специальная программа определяет номер дорожки, на которой находятся все основы, первые четыре буквы которых совпадают с указанными (конечно, возможен случай, когда этих букв меньше четырех, но это не меняет рассуждений). Далее, все основы, расположенные на ней, начиная с первой, сравниваются с переводимым словом до тех пор, пока не будет обнаружена та первая основа, которая либо совпала, либо целиком вложилась в переводимое слово. После этого вся статья, кроме самой основы, записывается в некоторую строку ДОЗУ. Важно отметить, что эта процедура выполняется в одну команду с помощью схемы, которая анализирует единицу, идущую сразу после основы, и сексту, содержащую сведения о длине статьи.

Выборка из армянского словаря производится аналогично. В начале армянской словарной статьи находятся две сексты, содержащие наименьший из номеров тех слов, которые синтезируются по этой статье. Третья секста содержит число, равное числу секст в остальной части статьи (основа и вся необходимая для синтеза конкретного слова информация). Выборка нужной статьи осуществляется сравнением номера синтезируемого слова с номерами статей, расположенных на дорожках барабанов одна непосредственно за другой (без локализации



адресов!) в порядке убывания номеров. При этом выбирается та статья, номер которой либо совпадает, либо является первым меньшим, чем номер синтезируемого слова. В ДОЗУ передается вся статья, но вместо номера статьи передается обратный код разности между номером синтезируемого слова и номером извлеченной статьи (последнее связано исключительно с техническим упрощением схемы выборки статьи)*.

§ 3. Операционное устройство

Как отмечалось выше, краткость программной реализации алгоритма перевода на машине «Гарни» достигается не только наличием специальных запоминающих устройств, но и рядом специфических операций, характерные особенности которых будут описаны в этом параграфе.

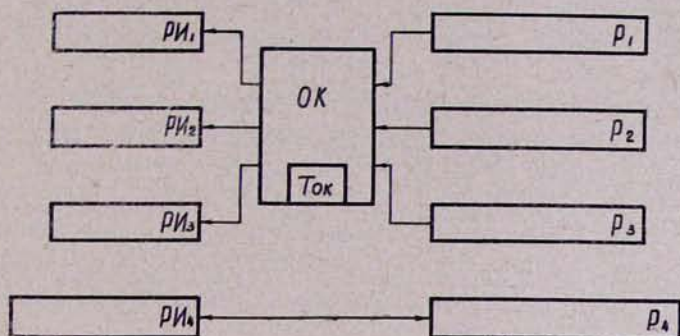


Рис. 5.

На рис. 5 изображен операционный узел (ОУ), который содержит три 32-разрядных регистра P_1, P_2, P_3 , один 36-разрядный регистр P_4 , три 8-разрядных регистра RI_1, RI_2, RI_3 и 14-разрядный регистр RI_4 . Ряд других регистров, нас в настоящее время не интересующих, на рисунке не указан. Выбор длины регистров P_1, P_2, P_3 основан на следующем соображении. Алгоритм в основном оперирует с простыми предложениями, длина которых редко превышает 32 слова. С другой стороны, для нужд переадресаций число разрядов в регистрах должно равняться некоторой степени двойки. Поэтому регистры машины «Гарни» содержат 32 разряда; создание регистров в 64 разряда привело бы к значительному увеличению весьма неэффективно используемого оборудования ОУ. Разрядность P_4 привязана к разрядности ячеек ПЗУ.

Регистры P_3, P_4 имеют двустороннюю связь с запоминающими устройствами машины, и на них выполняются основные логические и арифметические операции (см. таблицу команд машины «Гарни», приложение

* Подробнее об обращении к русскому и армянскому словарям см. статью Н. В. Вартаняна, Э. В. Егизаряна, Р. Л. Урутына «Организация словарей» в настоящем сборнике.



№ 2)*. Заметим, что из арифметических операций в машине «Гарни» предусмотрены только операции сложения и вычитания целых чисел, причем, если уменьшаемое меньше вычитаемого, то на регистр P_4 выдается не сама разность, а ее дополнение на этом регистре. В программах эти операции используются только для целей переадресации.

Регистры P_1, P_2 вместе с регистром P_3 служат для выполнения специальной группы операций, названных операциями «поиска конструкций». Как отмечалось выше, при реализации алгоритма синтаксиса мы оперируем в основном со шкалами, и при этом нас интересует, имеются ли во фразе слова с заданными признаками и в заданной последовательности. В рассмотренном в § 1 примере мы искали тройку слов, из которых первое обладает некоторым признаком α (прилагательное-существительное в именительном-винительном падежах), второе—признаком β (переходный глагол) и третье—признаком γ (существительное в именительном-винительном падежах), причем между крайними словами с этими свойствами нет других слов с теми же признаками.

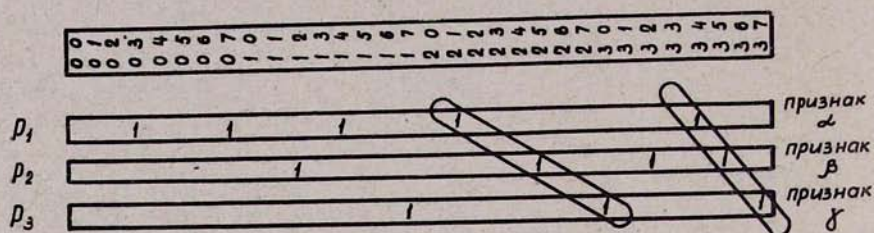


Рис. 6.

Пусть, например, шкалы признаков α, β, γ имеют вид, изображенный на рис. 6. Во фразе, которой соответствуют эти шкалы, имеются две тройки интересующих нас слов: тройка слов, имеющих номера 21, 25, 30 и тройка с номерами 34, 35, 37. Если шкалы признаков α, β, γ расположены в трех последовательных линиях ДОЗУ, то одной командой все они выдаются на регистры P_1, P_2, P_3 соответственно, затем содержимое этих регистров (разряд за разрядом) сдвигается налево, и старшие разряды регистров все время анализируются специальным автоматом, так называемым автоматом «определителя конструкций» (ОК)**. Как только автомат обнаружит искомую конструкцию (а в машине «Гарни» предусмотрена возможность обнаружения конструкций пяти различных типов), выполнение команды прекращается, номера обнаруженных компонент (в нашем случае числа 21, 25 и 30) передаются на регистры PI_1, PI_2, PI_3 соответственно, в специальный индикатор-триггер определителя конструкций (ТОК) передается единица, и машина переходит к выполнению следующей команды. При этом на регистре PI_1

* Полностью все команды описаны в статье Р. Л. Урутыяна «Система команд машины «Гарни» в настоящем сборнике.

** Подробнее см. в статье А. Г. Ованисяна и Р. Л. Урутыяна «Определитель конструкций» в настоящем сборнике.

запоминается номер последнего проанализированного разряда (тридцатого в нашем случае). Так как работа ОК начинается только с того разряда, номер которого записан на РИ₄, то достаточно повторить ту же команду, не изменив содержимого РИ₄, чтобы обнаружить вторую тройку. Поиск конструкции ведется до конца шкал, причем в специальных разрядах команды указывается, из какого числа отрезков состоит в данном случае анализируемые шкалы. После команды поиска конструкции ставится обычно команда, которая проверяет состояние ТОК и в зависимости от того, обнаружена или нет искомая конструкция, разветвляет дальнейший процесс. Легко видеть, что программная реализация такого поиска или, например, обнаружение во фразе двух слов с некоторыми свойствами, все слова между которыми обладают одним и тем же заданным свойством, заняли бы на универсальной машине много места и времени.

Характерной командой является также команда подсчета числа единиц в некоторой шкале. Ее содержательный смысл—сколько слов во фразе обладают данным свойством.

В § 1 уже были описаны операции обращения к словарям. Точно так же очень удобными для осуществления перевода являются команды ввода переводимого текста в литеры языка оригинала и вывода переведенного текста в литеры языка перевода. Поскольку надежно работающих читающих устройств в настоящее время нет, переводимый текст подлежит предварительной перепечатке на специальном «устройстве записи на магнитную ленту». Оно состоит из обычного телетайпа, снабженного, однако, специальной приставкой, с помощью которой весь текст, помимо записи его на бумагу (для визуального контроля работы машинистки), кодируется и записывается на магнитную ленту (МЛ).

При переписке должны быть опущены все точки, находящиеся внутри фразы (аббревиатуры типа «см. рис. 1» заменяются на «смотри рисунок 1»). Все формулы, знаки, числа и подобная небуквенная информация заменяется специальным знаком, после которого указывается объем замененного выражения (то есть число литер, которые должны быть пропущены при отпечатывании перевода для последующего вписывания от руки опущенного выражения). Если во время переписки текста обнаружена некоторая ошибка, то достаточно отпечатать специальный знак—«исправление ошибки», причем столько раз, сколько знаков пробела отпечатано после допущенной ошибки, и продолжать переписку текста, начиная с того слова, в котором была обнаружена ошибка. При вводе как это слово, так и все отпечатанные после него слова (до знаков «ошибка») будут исключены.

Устройство записи на магнитную ленту может использоваться независимо от самой машины. Это дает возможность вести переписку текста на нескольких устройствах одновременно, которые к тому же могут быть установлены непосредственно у заказчика, территориально не связанного с машиной.

Ввод текста осуществляется пофразно, причем каждое слово вводится в новую горизонтальную линию ДОЗУ, начиная с начала этой линии (машина анализирует код знака пробела). Ввод прекращается, когда поступает точка. Вся эта процедура осуществляется в одну команду, после которой машина переходит к выполнению следующей команды. Адрес линии, в которую вводится первое слово, указывается в команде.

Точно так же в одну команду осуществляется вывод переведенной фразы, слова которой (закодированные побуквенно) в процессе морфологического синтеза располагаются в последовательных линиях ДОЗУ. Начав с линии, указанной в команде, машина отпечатывает букву за буквой, пока не доходит до кода пробела, который сигнализирует необходимость перехода к следующей линии, и так продолжается до тех пор, пока не будет отпечатана точка.

Заметим, что как при вводе, так и во все время анализа и синтеза каждому знаку препинания и каждой формуле отводится отдельная линия ДОЗУ.

§ 4. Процесс перевода

Чтобы описать процесс перевода, надо предварительно рассмотреть блок-схему машины, изображенную на рис. 7, на котором указаны все имеющиеся связи между отдельными узлами машины. Помимо описанных в предыдущих параграфах узлов, на рис. 7 указаны устройства: управления (УУ)*, ввода вывода записи на магнитную ленту и пульт управления (ПУ).

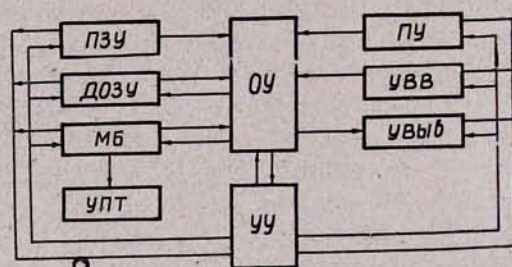


Рис. 7. Блок-схема машины «Гарни».

Устройство ввода с магнитной ленты (оно выполнено на лентопротяжном механизме МЭЗ—28) служит для пофразного ввода переводимого текста так, как это описано в предыдущем параграфе**. С этого же устройства вводятся в МБ оба словаря. Вывод как результатов промежуточных преобразований в литеры русского языка, так и окончатель-

* Подробнее об этом см. в статье Р. Л. Урутына «Синтез центрального устройства управления» в настоящем сборнике.

** Подробнее об этих устройствах см. в статье А. М. Арутюнова, Р. Б. Бардуги-меосяна «Устройство ввода и вывода» в настоящем сборнике.

ного перевода в литерях армянского языка осуществляется с помощью буквопечатающего телетайпа, входящего в УВыв. Кроме того, для осуществления небольших экспериментов с того же телетайпа можно от руки вводить текст в машину.

Для ввода словарей в машину они предварительно кодируются в восьмеричной системе, а затем с помощью того же устройства записи переносятся на магнитную ленту (при этом в начале каждой дорожки ставится номер дорожки, а в конце ее—специальный знак «конец дорожки»). Контроль правильности переписки на ленту осуществляется так же, как при переписке текстов. Магнитная лента устанавливается на УВВ, и очередная дорожка вводится сначала в верхнюю половину ДОЗУ (линии с номера 000 по 077), а затем в нижнюю (линии 100—177), после чего они сравниваются между собой. При несовпадении ввод повторяется, а при совпадении вся дорожка из верхней половины ДОЗУ передается на соответствующую дорожку МБ. Тут же дорожка считывается в нижнюю половину ДОЗУ, а затем две половины ДОЗУ вновь сравниваются между собой. В случае необходимости эта процедура повторяется до совпадения информации, введенной на дорожку МБ и считанной с нее, после чего машина переходит к вводу следующей дорожки. Вся описанная процедура осуществляется программно.

Если все же при переписке дорожки или при ручной кодировке будет допущена ошибка или возникнет необходимость внесения в словарь изменений или дополнений, то уже переписанная на ленту соответствующая дорожка вводится в ДОЗУ. В нем (по специальной программе) в нужном месте и на нужное число секст осуществляется сдвиг информации (что-то надо выкинуть), или наоборот—информация раздвигается (нужно что-то добавить), вносятся изменения и исправления (работой с пульта или программно) и затем исправленная дорожка переписывается из ДОЗУ на новую ленту. Эта запись из ДОЗУ на МЛ осуществляется с помощью устройства записи, которое подключается к каналам, передающим информацию из ДОЗУ на МБ, причем сами МБ на это время отключаются от машины. Конечно, значительно проще было бы хранить словари на перфокартах и на них же осуществлять все изменения, но мы предпочли описанный выше способ ввиду его значительно большей устойчивости и надежности в работе.

Для осуществления самого перевода магнитная лента с записанным на ней текстом устанавливается на УВВ. После пуска машины первая записанная фраза передается с ленты в верхнюю половину ДОЗУ (предполагается, что фраза содержит не более 64 слов), и машина переходит к осуществлению алгоритма перевода (см. рис. 1).

Для каждого слова из словаря извлекается соответствующая словарная статья, с помощью которой осуществляется морфологический анализ слов. Каждому знаку препинания и знаку формулы вырабатывается соответствующая информация и начинается синтаксический анализ. На первом этапе осуществляется членение фразы на простые предложения, после чего они передаются одно за другим в нижнюю по-

ловину ДОЗУ, где завершается их синтаксический анализ и осуществляется синтаксический синтез. Затем для каждого слова простого предложения осуществляется морфологический синтез, предложение отпечатывается, и программа переходит к следующему простому предложению. После того как будет переведена вся фраза, программа вновь запускает УВВ, в ДОЗУ передается следующая фраза, и весь процесс повторяется.

При необходимости сменить тематику переводимых текстов достаточно ввести в машину словари по соответствующей специальности. Нам представляется, что не только смена тематики, но даже смена языков может быть легко осуществлена на машине «Гарни» одной лишь заменой тех кассет ПЗУ, в которых хранятся программы, алгоритмы и лингвистические таблицы, связанные с данной парой языков. Если при этом алгоритмы переводов достаточно однотипны, то программы, реализующие алгоритм, могут быть оставлены без изменения.

Ք. Մ. ՏԵՐ-ՄԻՔԱՅԵԼՅԱՆ, Ռ. Լ. ՈՒՌՈՒՅԱՆ

ՔԱՐԳՄԱՆՈՒԹՅԱՆ ԱԼԳՈՐԻԹՄԻ ԿԱՌՈՒՅՎԱԾՔԸ
 ԵՎ «ԳԱՌՆԻ» ՄԵՔԵՆԱՅԻ ԸՆԴՀԱՆՈՒՐ ՆԿԱՐԱԳՐՈՒԹՅՈՒՆԸ

Ա մ փ ո փ ո լ մ

«Գառնի» մասնագիտացված մեքենայով թարգմանություններ կատարելիս կիրառվում է հատուկ ալգորիթմ, որի հիմնական էտապներն են նախադասությունների ձևաբանական և շարահյուսական անալիզն ու սինթեզը:

Անշուշտ, ալգորիթմի առանձնահատկությունները պահանջ են առաջադրել մեքենայի կառուցվածքին՝ որոշելով ռեգիստրների կարգային ցանցը, հիշողությունների ունակությունն ու նրանց օգտագործման եղանակները: Ներմուծման և արտածման սարքերը, Հարց է ծագում՝ ստեղծել սկզբունքորեն նոր տիպի հանգույցներ, որոնք բացակայում են համընդհանուր հաշվիչ մեքենաների մոտ: Այսպես օրինակ՝ կառուցվածքների որոշիչը հնարավորություն է տալիս պարզելու այս կամ այն քերականական հատկանիշ ունեցող բառերի հաջորդականության գոյությունը նախադասության մեջ և այլն:

Համառոտակի նկարագրենք մեկ նախադասության թարգմանությունը: այն մեքենային մատուցվում է ձեռքով՝ տպագրական մեքենայի կամ ավտոմատ՝ մագնիսական ժապավենի օգնությամբ:

Ձևաբանական վերլուծության ընթացքում մեքենայական շառարաններից հիշողության տարբեր բջիջներում գրանցվում են նախադասության յուրաքանչյուր անդամին վերաբերվող բառարանային հոդվածներ:

Այնուհետև կատարվում է նախադասության շարահյուսական վերլուծություն, որի օգնությամբ վերականգնվում է նախադասության անդամների քերականական միարժեքությունը: Ընտրելով համապատասխան բառերը իրենց հոդվածներով՝ շարահյուսական ու ձևաբանական սինթեզի ընթացքում կառուցվում է հայերեն նախադասությունը, որն արտածվում է տպագրական մեքենայի օգնությամբ:

Л И Т Е Р А Т У Р А

1. Р. А. Базмаджян, Т. К. Жанполадян, Т. М. Тер-Микаэлян. Об одном варианте морфологического анализа и методе его реализации на специализированной машине. Сборник «Проблемы кибернетики», вып. 15, М., «Наука», 1965.
2. И. С. Дургарян, Т. М. Тер-Микаэлян. Некоторые соображения о структуре специализированной цифровой машины для языкового перевода, НТИ, М., 1964, № 2.
3. В. М. Григорян, Т. М. Тер-Микаэлян, Р. Л. Урутян. Алгоритмы русско-армянского машинного перевода и его реализация на специализированной машине. Тезисы докладов на конференции по МП, Ереван, 1967.