

Особенности письменного машинного перевода

При анализе текста применительно к некоторому данному языку прежде всего возникает вопрос о том, с каким отрезком мы имеем дело и на какие единицы он последовательно разбивается. Существенно определить, какой отрезок переводимого текста содержит грамматические связи, осуществляющиеся лишь внутри него и не переходящие за его границы. Определение границ такого отрезка не столь просто. Известную помощь оказывает пунктуация, но не во всех случаях. Практически берется отрезок, содержащийся между двумя точками, т. е. тот, который синтаксически представляет собой предложение, возможно, так называемое сложное предложение. Большая часть грамматических связей действительно заключается внутри него. Но, во-первых, внутри этого отрезка могут находиться различные отрезки, представляющие собой некоторые единицы, никаких грамматических связей между которыми нет. Это относится, прежде всего, к последовательности предложений, не связанных между собой никакими грамматическими средствами (союзами, относительными словами) и лишь разделенных на письме запятыми. Во-вторых, некоторые связи выходят все же за пределы отрезка, графически ограниченного точками. Это относится к местоимению типа *он* (*она*, *оно*), выбор форм которого в ряде языков может быть обусловлен принадлежностью к определенной категории (рода, лица и т. п.) существительного, лежащего за пределами данного отрезка.

Что касается единиц, расположенных внутри отрезка, ограниченного точками, то их выделение в некоторых случаях также осуществляется при помощи пунктуации, иногда же пунктуация оказать помощи не может, что в определенных случаях зависит от принятой для данного языка пунктуационной системы. Так, например, из соответствующих друг другу по значению французского оборота с относительным местоимением *qui* и русского с *который* русский оборот выделяется запятыми, выступающими как внешние показатели его границ. Французский оборот запятыми не выделяется, начальная его граница определяется формой *qui*, конечная же должна быть определена другими способами (не при помощи пунктуации и не при помощи особого слова, подобно тому как определяется начальная граница).

Нижеследующие статьи посвящены описанию некоторых специальных вопросов машинного перевода и работ по его реализации.

П. С. Кузнецов, А. А. Ляпунов, А. А. Реформатский

МАШИННЫЙ ПЕРЕВОД С ФРАНЦУЗСКОГО ЯЗЫКА НА РУССКИЙ

В настоящей работе излагаются принципы построения словаря и «грамматики» (свода специальных правил) для машинного перевода с французского языка на русский научного текста (из области математики), а также описывается сам процесс перевода.

Для перевода необходимо иметь следующие данные: 1) «Словарь основ»; 2) «Словарь оборотов»; 3) «Таблицы перевода предлогов»; 4) «Правила различения омонимов»; 5) «Таблицы окончаний» (французских и русских); 6) «Группа анализирующих правил»; 7) «Группа синтезирующих правил». По мере надобности эти данные вводятся в «память» машины.

I

Для машинного перевода математических текстов с французского языка на русский нами был подготовлен особый словарь на основе анализа словарного состава различных математических текстов. Подсчитывались как слова, так и словосочетания. Было рассмотрено 20 500 слов, среди них оказалось 2300 различных. Из этих слов были отобраны те, которые встретились не менее четырех раз; их оказалось около 1000. К этой тысяче слов были добавлены без статистических подсчетов термины, необходимость которых была очевидна (их около 50), а также все служебные слова. В результате получился словарь приблизительно на 1400 слов.

Особенность этого словаря в том, что в нем помещены не слова, а основы. Под термином «основа» здесь понимается часть слова, которая остается графически неизменной во всех его формах, т. е. не совсем то, что обозначает этот термин в обычной грамматике. Часть слова, оставшуюся после отделения «основы», мы называем «окончанием». Так, для слова *travail* в нашем словаре основой будет *trava-* [так как у этого слова две формы: *trava-il* (ед. число) и *trava-ux* (мн. число)], для *parler* основа *parl-*, для *finir* основа *fini-* и т. д. Отметим, что для большинства французских существительных «основа» в нашем понимании совпадает с формой единственного числа.

поэтому почти все существительные приводятся в той же форме, что и в обычных словарях. Все неизменяемые слова приводятся целиком.

Несколько сложнее обстоит дело с глаголами. У многих из них после выделения графически неизменной части («основы») получается слишком большое разнообразие «окончаний»:

<i>v-enir</i>	<i>m-ener</i>
<i>v-ient</i>	<i>m-ène</i>
<i>v-iennent</i>	<i>m-enons</i>

т. е. получаются новые «окончания» для инфинитива (*-ener*, *-enir*), для 3-го лица ед. числа (*-ient*, *-ène*), для 3-го лица мн. числа и т. д. Поэтому мы прежде всего отделили в глаголах стандартные (т. е. общие для основных групп) окончания. Оставшаяся часть слова не могла быть принята как «основа» в нашем понимании, так как она графически изменяема. Поэтому различные для различных форм варианты основы глагола помещены в словаре как отдельные слова-«основы», каждое со своей полной словарной статьей. Например, глагол *venir* дается в следующем виде:

<i>cen-</i>	<i>ons, -u)</i>
<i>vien-</i>	<i>(-s, -t)</i>
<i>vienn-</i>	<i>(-ent)</i>
<i>viend-</i>	<i>(-ra, -rait)</i>

Относительно расположения основ в словаре надо заметить следующее: для упорядочения букв принят алфавитный порядок. Необычным является только то, что короткая основа в нашем словаре стоит позади всех более длинных, в которых она целиком содержится, т. е. *suppos(-er)* помещено после *supposition*, *-généra(-l)* после *généralisation* и *généralis(-er)* и т. д. Такое расположение оказывается более удобным при поиске.

Каждая основа в словаре снабжена словарной статьей, которая содержит: 1) переводы; 2) информацию французскую: а) указание на часть речи, б) указание на обороты, в) предложный код, г) грамматические характеристики; 3) информацию русскую (к переводящей основе или к каждой из них, если их несколько): а) указания о выборе основы перевода, б) грамматические характеристики; 4) помету о выборе 2-х основ (т. е. указание на то, что для данного слова нужно выбрать два родственных варианта основы переводящего слова). Отметим, что пункты 1, 2, 3 обязательны для всех слов-основ, а пункт 4 важен только для некоторых.

Перевод

Для подавляющего большинства французских слов перевод состоит из одной или нескольких основ одного и того же переводящего слова. Несколько основ даются в том случае, если изменения переводящего слова отражаются в графических изменениях (чередованиях) основы. Например:

<i>équation-</i>	— «уравнени-» (одна основа)
<i>nombre-</i>	— «числ-» (две основы)
	«чисел-»
<i>suppos(-er)</i>	— «предполага-» (-ть)
	«предполож-» (-ить, -им, -енный)
	«предположи-» (-л).

В некоторых случаях для французского слова дается два переводящих слова (слов с двумя переводами оказывается около 50), причем выбор одного из них формально никак не обусловлен и зависит только от смысла фразы. В таких случаях в соответствующей русской фразе будут приведены оба варианта перевода слова, так как пока решено не прибегать к анализу смысловых связей слов и использовать при переводе только их формальные отношения.

Здесь машина окажется в положении переводчика, не знающего той области математики, к которой принадлежит переводимый текст. Переводчик-неспециалист не всегда может выбрать правильный вариант переводящего слова.

формация французская

а) Указание на часть речи. Каждое слово-основа в словаре снабжено пометой, указывающей, к какой части речи оно относится (об омонимах типа *point — le point*, *entre — il entre*, относящихся к разным частям речи, см. ниже). Это оказалось возможным потому, что у большинства французских слов (по крайней мере в лексике математического текста) принадлежность к той или иной части речи выражена формально (синтетическими средствами) в достаточной степени, так что эту принадлежность можно определить по отдельному взятому вне контекста слову.

б) Указание на обороты. При переводе французского текста часто бывает необходимо переводить не отдельные слова, а целые словосочетания, которые мы в дальнейшем будем называть «оборотами».

Под оборотами мы понимаем все те комбинации из нескольких слов, буквальный перевод которых одного за другим (по данным нашего словаря), без учета формы их сочетания, привел бы к искажению смысла. Такие обороты следует переводить как единое целое. Так, *à sa gauche* нельзя перевести дословно «в-его, ее-левый (-ую)», а надо «налево от него, нее».

Все обороты помещены в специальный «словарь оборотов», где каждый из них имеет словарную статью, аналогичную словарным статьям в «словаре основ». Обороты делятся на «целые» и «нецелые».

Мы называем оборот *целым*, если все входящие в него слова в переводимой фразе обязательно стоят подряд и между ними нельзя вставить никакого другого слова (например, *de plus en plus, en effet, à présent*). В *нецелые* обороты могут быть введены некоторые другие слова, причем для каждого такого оборота точно указывается, какие именно слова или части речи и на какое место могут вставляться (например, *aussi + adjectif + que* «*столь же + прилагательное + как и*», *à + numéral, substantif + près* «*с точностью до + числительное, существительное*»). В словарной статье оборота имеется обязательная помета, является ли он *целым* или *нецелым*.

В обороте выделяется основное значащее слово, и обороты в «словаре оборотов» располагаются в алфавитном порядке основных слов, а в пределах одного слова (которое может быть основным в нескольких оборотах) — по количеству слов в обороте, причем короткие обороты ставятся после более длинных. Так, *au plus* «по большей мере» стоит после *de plus en plus* «все больше», а *sous forme* «в виде» после *sous la forme* «в виде». Обороты в «словаре оборотов» последовательно пронумерованы. Указание на оборот имеется у слов, могущих быть основными словами того или иного оборота, и состоит оно из 2 частей: 1) порядковый номер *первого* оборота (в который данное слово входит как основное) в «словаре оборотов»; 2) число оборотов, в которые данное слово входит как основное. (Об использовании этих данных при переводе оборотов смотри стр. 115.)

в) *Предложный код*. Предлоги как часть речи при переводе допускают наибольшее число вариантов, причем падеж, которого требует предлог, находится в зависимости от того, какой русский предлог выбран. Некоторые предлоги всегда переводятся однозначно: *contre* «против», *vers* «к», *jusqu'à* «до», *autour de* «вокруг» и т. д. Семь предлогов (*à, dans, de, en, par, pour, sur*) переводятся в зависимости от того, каким словом они управляются (чаще всего — глаголом, но иногда — существительным и прилагательным). Для перевода предлогов составлены специальные «Таблицы перевода предлогов». В каждой из них дается список французских предлогов, которые могут управляться тем или иным *одним* словом, и каждому предлогу приписывается один перевод, в соответствии с переводом управляющего слова; при этом указывается, какого падежа требует переводящий предлог.

«Предложный код слова» — это номер той из «Таблиц перевода предлогов», в которой помещены предлоги, управляемые данным словом, и в которой им даны именно те переводы, какие требуются в сочетании их с данным словом. Например, основа *appliqu-* (глагол *appliquer* «применять») имеет в словаре код «16». «Таблица перевода предлогов» № 16 выглядит так:

<i>à</i>	к	дательный падеж
<i>dans</i>	в	предложный падеж
<i>en</i>	е	предложный падеж
<i>par</i>	(без предлога)	творительный падеж
<i>pour</i>	для	родительный падеж
<i>sur</i>	на	винительный падеж

Это означает, что *appliquer à* надо переводить «применять к чему-либо», *appliquer dans* или *en* — «применять в чем-либо» и т. д.

Предложный код имеют почти все глаголы и многие прилагательные (например, *indépendant de, parallèle à* и т. д.), а также некоторые существительные (*théorème sur*).

Если несколько слов управляют одними и теми же предлогами, причем переводятся эти предлоги одинаково при переводах всех этих управляющих слов, то всем этим словам приписывается общий предложный код.

г) Грамматические характеристики сводятся к следующим указаниям: для существительного — род, тип образования множественного числа; для глагола — переходный или непереходный, спрягается с *avoir* или с *être*, номер группы (об этом смотри стр. 117).

Для других частей речи имеются указания менее общего характера. Так, у пяти союзов — *et, mais, ni — ni, ou, c'est-à-dire* имеется помета «сочинительные» (они могут соединять однородные члены предложения); все остальные союзы (имеющиеся в нашем словаре) считаются подчинительными.

У предлогов *entre, parmi, avec* имеется указание, что при однородных членах предложения после сочинительного союза они не повторяются (т. е. *entre la maison et l'arbre*, а не *entre la maison et entre l'arbre*), тогда как, например, *à* повторяется (*à l'école et à la maison*).

Информация русская

а) Указание о выборе основы. Как уже говорилось, перевод дается в виде нескольких основ, если в переводящем слове происходят какие-либо чередования. В этом случае «Указание о выборе основы» представляет собой номер соответствующего правила из группы «синтезирующих правил». Это правило устанавливает, какую из приведенных в словаре переводящих основ следует выбрать для образования той или иной формы данного слова.

б) Грамматические характеристики сводятся к следующим указаниям: для существительного — «указание о выборе окончаний» (т. е. номер таблицы соответствующих данному существительному падежных окончаний), род; для глагола — «указание о выборе окончаний» (т. е. номера таблиц окончаний настоящего времени, будущего времени, прошедшего времени, неопределенной формы, деепричастий настоящего и прошедшего времени); «указание о выборе суффикса причастий».

Помета о выборе двух основ

Эта помета имеется только при некоторых основах и состоит из двух элементов: указание на то, что одновременно с данной основой следует выбрать из словаря еще одну; указание, какую именно вторую основу следует выбрать (объяснение этому — см. ниже, стр. 115).

Итак, здесь описано содержание и строение одной словарной статьи. Но в нашем словаре есть основы, имеющие при себе две словарные статьи: это случаи омонимии. Основа *fin-* (прилагательное «тонкий» и существительное «конец») имеет две совершенно различные словарные статьи. То же самое у основ *ensemble-* («множество» и «вместе»), *point-* («точка» и отрицательная частица) и т. д. Наткнувшись в тексте на слово с подобной основой, машина выбирает из словаря основу с обеими словарными статьями. После окончания поиска в словаре и обработки оборотов машина, пользуясь «Правилами различения омонимов» (см. стр. 115), «решит», какая именно из двух статей нужна в данном случае.

Такова в общих чертах структура «Словаря основ».

II

Процесс машинного перевода состоит из следующих операций: 1) поиск слова в словаре и перенос его вместе с его словарной статьёй в рабочую часть «памяти» машины; 2) обработка оборотов; 3) различение омонимов; 4) «анализ» — обработка переводимых слов в определенной последовательности (по частям речи) при помощи группы анализирующих правил; 5) «синтез» — построение русской фразы при помощи группы синтезирующих правил на основе сведений, полученных посредством анализирующих правил.

Поиск слов в словаре

Перевод начинается с того, что машина берет из переводимого текста одно слово за другим и ищет для каждого из них в словаре соответствующую основу. Задача поиска состоит в том, чтобы, сравнивая искомое слово с основами, данными в словаре, найти максимальную по числу букв основу, целиком уместящуюся в переводимом слове.

Благодаря принятому нами расположению основ в словаре первая по порядку основа, целиком уместившаяся в переводимом слове, будет искомая. Действительно, любая другая основа, тоже целиком содержащаяся в переводимом слове, как более короткая, будет стоять в словаре позже.

Например: при переводе слова *généralisons* находим в словаре основы: 1) *généralisation-*, 2) *généralis-*, 3) *généra-*. Первая не укладывается в переводимом слове, вторая укладывается в нем полностью; она и будет искомой, так как третья, которая тоже укладывается полностью в слове, имеет меньше совпадающих букв (6 против 9

у *généralis*-). Найденная основа передается из словаря в рабочую часть «памяти» машины вместе с ее словарной статьей (статьями).

В дальнейшем, когда идет речь о «выборе основы из словаря», всегда подразумевается именно этот перенос основы с ее словарной статьей (статьями). Если была отметка о выборе второй основы, то в рабочую часть «памяти» машины также переводится вторая основа с ее словарной статьей.

Обработка оборотов

Когда ко всем словам фразы¹ выбраны из словаря соответствующие основы со словарными статьями, машина проверяет, нет ли в какой-либо из этих статей «указания на обороты». Если указание отсутствует, можно переходить к следующему этапу перевода («различение омонимов»). Если такое указание имеется, машина обращается к «словарю оборотов» и по номеру, данному в указании, находит первый оборот, в котором снабженное указанием слово является основным. Затем машина сравнивает этот оборот с переводимой фразой. Если он полностью уложится во фразу (т. е. в составе фразы имеются в ее слова, входящие в оборот, и они при этом расположены относительно друг друга и относительно основного слова совершенно так же, как в самом обороте), то оборот считается найденным и словарные статьи всех входящих в него слов заменяются словарной статьей оборота. Если оборот не уложился, то машина последовательно сравнивает с переводимой фразой все обороты с данным основным словом (число оборотов указано при основном слове). Если же ни один из оборотов, которые относятся к слову, имеющему «указание на обороты», «не уложился» в переводимой фразе, то это значит, что в данном случае слово не входит в оборот и его следует переводить самостоятельно.

Различение омонимов

Под «различением омонимов» понимаются две следующие операции:

I. Выбор одной из двух основ, полученных одновременно по специальной помете (см. стр. 114).

II. Выбор одной из нескольких статей, имеющих при данной основе (см. стр. 114). Разберем подробнее каждую операцию.

I. Во французском языке имеется ряд слов, которые могут быть и формой глагола, и существительным, например: *forme* «форма» и *(il) forme* «(он) образует», *fait* «факт» и *(il) fait* «(он) делает», *limite* «предел» и *(il) limite* «(он) ограничивает». В нашем словаре для таких слов имеются две разные основы: *forme-* (сущ.), *form-* (глагол) или *fait-* (сущ.), *fai-* (глагол). Но, действуя по сформулированному выше принципу поиска (принцип максимального совпадения), машина, отыскивая основу для глагола *(il) forme*, *former*, *formera*, найдет основу существительного, как более длинную. В таких случаях у основы существительного имеется в словаре помета, что вместе с ней надо выбрать из словаря основу глагола, и указывается, какую именно. Чтобы определить, какая основа нужна для перевода, переводимое слово сначала сравнивается с основой существительного и рассматривается окончание, которое при этом остается. Имеется два случая²: или остается какое-либо окончание, или же окончание отсутствует. Рассмотрим первый случай.

А) Если осталось окончание *-s*, то рассматриваемое слово является существительным и для его обработки берется основа существительного, а основа глагола отбрасывается. (Окончание *-s* может быть не только окончанием множественного числа существительного, но и окончанием 2-го лица единственного числа глагола, однако вторую возможность мы отбрасываем, так как в математических текстах глаголы во втором лице единственного числа не встречаются.)

Б) Если остается любое другое окончание, то рассматриваемое слово — глагол, и для его перевода берется основа глагола, основа существительного отбрасывается.

Во втором случае (т. е. когда не осталось никакого окончания) рассматриваемое слово является либо существительным в единственном числе, либо глаголом в 3-м—4-м³ лице ед. числа.

Эти два случая различаются следующим образом [омоформу «существительное-глагол» типа *(une) forme* — *(il) forme* будем условно обозначать VS]:

А) VS, перед которым стоят детерминативы⁴, к р о м е *le*, *la*, или любой предлог,

¹ Фразой мы считаем отрезок текста между двумя точками или между точкой и точкой с запятой.

² Отметим, что как «правила различения омонимов», так и «анализирующие правила» построены в основном по принципу дихотомического деления, т. е. на каждом данном этапе перед машиной ставится вопрос о выборе из двух возможностей.

³ Так называемое «четвертое лицо» — условное обозначение форманта *-e* (совпадающего окончания 1-го и 3-го лица ед. числа глаголов I группы в *présent*).

⁴ Детерминативами называются артикли, указательные и притяжательные местоимения.

кроме *en*, является существительным. (Оговорку о *le*, *la* и *en* пришлось сделать потому, что *le* и *la* могут быть не только детерминативом, но и прямым объектным местоимением при глаголе, например: *il le limite*, *il la forme* и т. д.; *en* может быть не только предлогом, но и приглагольным местоимением, например: *il en forme quelque chose*.)

Б) VS, перед которым стоят *le*, *la* или *en*.

1. В конструкции:

$$S_{NA} \text{ (или } Pr_N) + \overset{le}{\underset{en}{+}} la + VS$$

а) При отсутствии запятой или сочинительного союза перед *le*, *la*, *en* — VS является глаголом.

б) При наличии запятой или сочинительного союза в вышеупомянутой конструкции [т. е. $S_{NA} \text{ (или } Pr_N) + \overset{le}{\underset{en}{+}} la + VS$]

а) VS является глаголом, если за ним не следует глагол до конца фразы или до союза;

б) VS является существительным, если за ним следует глагол.

2. Если VS, следующее после *le*, *la*, *en*, не входит в упомянутую конструкцию, то это существительное.

В) В остальных случаях VS является глаголом.

Подчеркнем, что в этом и в последующих случаях S может означать как отдельное существительное, так и сочетание существительного с прилагательными и наречиями, которые могут стоять как слева, так и справа от него.

Кроме омоформ «существительное-глагол» имеется единственный случай омоформы «предлог-глагол»: *entre* «между» и (*il*) *entre* «(он) входит». Для этого случая выбор нужной основы производится по следующему правилу: *entre* перед группой «детерминатив+существительное» — предлог, в остальных случаях — глагол.

II. В нашем словаре имеются различные случаи омонимии, когда основа имеет при себе не одну словарную статью. Например: 1) *fin* «тонкий» или «конец», *plan* «плоский» или «плоскость» и т. д., то-есть омонимия прилагательного и существительного; 2) *pas*-, *point*- «шаг», «точка» или отрицательные частицы; 3) *soit* «пусть» или форма 3-го лица ед. числа *subjonctif* от глагола *être*; 4) *ensemble*- «множество» или «вместе»; 5) *même*- «тот же» или «даже»; 6) *plus*-, *moins*- «больше», «меньше» или «плюс», «минус»; 7) *suivant*- «следующий» или «по» и т. д.

Для первого случая формулируется правило различения не отдельных слов, а лексико-грамматических категорий — существительного и прилагательного (случай омоформы «существительное-прилагательное» будем условно обозначать SA).

Правило различения в этом случае таково: SA после детерминатива или предлога при отсутствии справа существительного, которое может быть отделено прилагательными, наречиями и сочинительными союзами, является существительным. В остальных случаях SA — прилагательное.

Для всех других случаев формулируются правила применительно к отдельным словам, например:

С л у ч а й 2: *pas*, *point* являются отрицательными частицами, если они стоят непосредственно после глагола (при этом они могут отделяться только наречиями) и в конструкции типа *ne + pas + infinitif*, в остальных же случаях эти слова — существительные.

С л у ч а й 4: *ensemble* после детерминатива или предлога (при этом может отделяться прилагательными, наречиями и сочинительными союзами) существительное, в остальных случаях — наречие. Аналогичные правила имеются для всех случаев омонимии, встречающихся в нашем словаре.

Кроме описанных двух случаев различения омонимов (выбор одной из двух основ и выбор одной из двух словарных статей), встречается еще один случай: некоторые французские глаголы в контексте могут быть как переходными, так и непереходными, например: *changer* «менять» и «меняться»; *continuer* «продолжать» и «продолжаться» и др. Такие глаголы (обозначаемые V_{ti}) считаются переходными в конструкции $S_{NA} + V_{ti} + S_{NA}$ при отсутствии запятой перед вторым S_{NA} . В эту конструкцию может входить еще и существительное с предлогом (между V_{ti} и вторым S_{NA}).

«Правила различения омонимов» неприменимы к словам, которые имеют несколько переводов, различия только по смыслу. В этих случаях машина будет выдавать все имеющиеся у нее варианты перевода. Поскольку различные варианты перевода принадлежат к одной и той же части речи, то они, следовательно, требуют установле-

¹ S_{NA} — условное обозначение существительного без предлога, т. е. в *nominatif* или *accusatif*; Pr_N — условное обозначение местоимения в *nominatif*.

нии одних и тех же данных (число и падеж — для существительных; время, вид, число, лицо или род — для глагола и т. д.), т. е. обработка таких слов анализирующими правилами ничем не отличается от обработки слов, имеющих одну словарную статью с одним переводом. Поэтому можно считать, что анализирующие правила после обработки оборотов и различения омонимов имеют дело со словами, каждому из которых в словаре соответствует одна основа с одной словарной статьей.

«Анализ»

Под «обработкой слова анализирующими правилами» понимается анализ форм французского слова и его положения во фразе среди других слов в целях определения форм и места соответствующего русского слова для выражения соответствующих связей между словами. «Анализирующие правила» сгруппированы по частям речи. В рабочем участке «памяти» машины перед обработкой «анализирующими правилами» имеются: 1) французские слова, разделенные на словарные основы и на окончания, которые получены вычитанием словарных основ из переводимых слов; 2) по одной словарной статье для каждого из слов; 3) свод правил обработки очередной части речи; 4) таблица французских окончаний.

Последовательность обработки такова: глаголы — частицы — предлоги — существительные — местоимения — прилагательные и причастия, играющие роль прилагательных. Последовательность эта не случайна. Ведь и при обычном переводе французского текста часто начинают с глагола не только потому, что он может быть смысловым ключом фразы, но и потому, что глагол — наиболее «синтетическая» часть речи во французском языке, он сохранил наибольшее количество форм (многие из них существуют лишь графически, но для машины только это и нужно). Французский глагол в своих формах выражает лицо, число, время, наклонение, возвратность; во многих случаях глагол можно перевести, исходя только из его форм, не обращаясь к связям с другими словами. Французское же существительное вовсе не выражает падежа, и падеж переводящего русского существительного обычно зависит от перевода глагола, управляющего существительным непосредственно или посредством предлогов. Поэтому существительное удобнее обрабатывать после глаголов и предлогов, а прилагательные — после существительных, так как они относятся к существительным и согласуются с ними. В качестве примера «анализирующих правил» здесь рассматриваются «Правила обработки глагола».

Форма русского глагола определяется следующими морфологическими данными: 1) время, вид; 2) число; 3) лицо (в настоящем и будущем); 4) род (в прош. вр.); 5) возвратность; 6) наличие отрицания. Определение наклонения русского глагола для машины не имеет смысла, так как в русском языке глагол в своих грамматических формах наклонения не выражает: *Я сделал бы, если бы мог* — в этой фразе с точки зрения машинного перевода представлены глаголы прошедшего времени, совершенного и несовершенного вида, мужского рода, невозвратные и неотрицательные. «Бы» является для машины отдельным словом, которое ставится иногда после глагола, например при переводе: *Si je le faisais* «если я делал бы это». «Правила обработки глагола» служат для определения всех перечисленных выше данных и исходят при этом, во-первых, из морфологической характеристики французского глагола и, во-вторых, из его отношений с другими словами.

Для некоторых «особых» (для машины) глаголов, таких, как, например, *être, aller*, отдельные формы приведены в словаре в качестве самостоятельных слов. При этих формах, кроме обычной словарной статьи, указаны время, лицо, число и иногда еще помета «окончателъности», которую поясним ниже.

Обработка глагола распадается на две операции: 1) работа с «таблицами окончаний» (получение предварительных данных) и 2) работа с правилами (уточнение предварительных данных). Если глагол «особый», то его предварительные данные (время, вид, лицо, число, «окончателъность», вставка «бы») получают непосредственно из словаря. Если глагол не «особый», то сначала по таблицам окончаний определяют предварительные данные. Для этого окончание, выделенное путем вычитания основы, сверяется с соответствующей данному глаголу таблицей окончаний. Таких таблиц 12, по числу различных глагольных групп (I спряжение, II спряжение и 10 групп архаического спряжения). У каждого глагола в словаре указан номер его группы, т. е. номер нужной таблицы (см. стр. 114). Окончания в таблицах расположены не по алфавиту, а по частотности, сначала *présent ind.* — 3-е лицо ед. числа (в I группе 4-е лицо ед. числа), 3-е лицо мн. числа, 1-е лицо мн. числа, 1-е лицо ед. числа; затем — *infinitif, participe passé, participe présent, futur, conditionnel, imparfait subjonctif* (порядок приведения личных окончаний — как в *présent ind.*).

Проверяемое окончание сравнивается с окончаниями таблицы, пока не получим полного совпадения. При каждом окончании французского глагола в таблице дается указание на соответствующее время, вид, число, лицо (для личных форм) переводящего глагола. Это оказалось возможным потому, что в общем формы времени русского глагола обычно имеют соответствия в системе времен французского языка: *présent*

ind. переводится обычно настоящим временем, *imparfait* — прошедшим несовершенного вида, *passé composé* — прошедшим совершенного вида и т. д.

Кроме этих данных, в таблице есть еще помета «окончателности» — указание, что полученные данные в уточнении не нуждаются (т. е. вторая операция отпадает).

После того как получены предварительные данные (из таблицы, а для особых глаголов — прямо из словаря), они уточняются при помощи правил, которые приведены ниже. Сначала выясняется, какое грамматическое время получено по предварительным данным, и в зависимости от этого приводится в действие одно из правил №№ 1, 2, 4, 5, 6. В каждом из этих правил указано, к какому следует перейти после его выполнения.

П р а в и л о 1. Предположим, что по предварительным данным получено «настоящее время».

а) Если перед глаголом в 1-м лице мн. числа нет *nous* (которое может быть отделено от глагола наречиями и местоимениями), настоящее время следует заменить будущим совершенного вида. Здесь имеется в виду форма 1-го лица мн. числа в *impératif*: *considérons la série...*¹ «рассмотрим ряд...»

б) Если перед глаголом есть *si*, отделенное от него местоимением в именительном падеже или существительным без предлога и наречиями, а в пределах фразы есть другой глагол в *futur*, то настоящее время надо заменить на будущее совершенного вида. Здесь имеется в виду условное придаточное, вводимое *si*: *si l'on développe, suivant les puissances de Z l'expression..., on verra que le coefficient...* (стр. 86) «если мы разложим по степеням *Z* выражение..., мы увидим, что коэффициент...»

в) Если перед глаголом имеется *que*, отделенное местоимениями и наречиями, то уточнение следует отложить до перевода *que*; если *que* переводится как союз «чтобы», то настоящее время надо заменить прошедшим совершенного вида. Здесь имеется в виду совпадение *subjonctif* и *indicatif* в I группе. Так как машина не различает графически совпадающие формы, то для нее *parle* в (*nous savons*) *qu'il parle* и (*nous voulons*) *qu'il parle* оказывается одной и той же формой, поскольку *-e* помечено в таблице как окончание настоящего времени 4-го лица ед. числа. Признаком того, что форма с окончанием *-e* — *subjonctif*, является наличие в русском переводе союза «чтобы». Обычно *subjonctif* переводится прошедшим временем совершенного вида + «бы», причем при наличии слева союза «чтобы» снимается «бы».

Выбор совершенного вида в этом и в некоторых других случаях обусловлен причинами практического удобства. В подавляющем большинстве случаев перевод будет верным, искажения смысла не произойдет, а это основное, что нас интересует, так как некоторая корявость части фраз в первом варианте машинного перевода неизбежна.

г) При отсутствии условий, указанных в первых трех пунктах, следует оставить настоящее время. После правила 1 переходим к правилу 2.

П р а в и л о 2. Предположим что по предварительным данным получено «будущее время совершенного вида».

а) Если имеется помета окончателности, то уточнения не требуется. Помета окончателности ставится при окончаниях *futur*, так как *futur* всегда переводится будущим временем.

б) Если пометы окончателности нет и в пределах фразы имеется союз *si* «если», то будущее время следует заменить прошедшим временем несовершенного вида + «бы».

в) Если нет ни пометы окончателности, ни *si* «если», то следует оставить предварительные данные.

«Будущее время» без пометы окончателности стоит в таблицах при окончаниях *conditionnel*; случай (б) имеет в виду *conditionnel* в главном предложении условного периода, случай (в) имеет в виду модальное употребление *Conditionnel* в самостоятельных предложениях, где он будет переводиться будущим временем. Такой перевод оказался практически удобным. Например, *la démonstration s'appliquerait sans modification à une fonction telle que...* (стр. 89) «доказательство применится без изменения к такой функции, что...» После правила 2 переходим к правилу 3.

П р а в и л о 3. Определение лица. В случаях, когда окончания 1-го и 3-го лица французских глаголов совпадают, этим окончаниям соответствует в таблицах условное 4-е лицо. Русскому прошедшему времени и неопределенной форме, не нуждающимся в определении лица, в таблице дается условное «нулевое» лицо.

Правило уточнения лица состоит в следующем: 4-е лицо при наличии слева местоимения *je*, которое может быть отделено местоимениями и наречиями, следует переводить 1-м, а при отсутствии этого местоимения — 3-м лицом. Для остальных лиц оставить предварительные данные. После правила 3 переходим к правилу 7.

П р а в и л о 4. Предположим, что по предварительным данным получено «прошедшее время».

а) Если глагол — несовершенного вида и перед ним имеется *si* «если», то следует

¹ См. Е. Borel, *Méthodes et problèmes de théorie des fonctions*, Paris, 1922, стр. 71. (В дальнейшем указания на страницы этой работы даются в тексте.)

заменить прошедшим временем совершенного вида и после союза «если» добавить частицу «бы». Если же *si* отсутствует, то оставить по предварительным данным. Прошедшее время несовершенного вида дается в таблицах для окончаний *imparfait*. Уточнение с *si* имеет в виду ирреальный условный период.

б) Если в переводимом тексте было «несовершенный вид + бы», а перед глаголом имелось *que*, то уточнение возможно только после перевода *que*; если *que* переводится как союз «чтобы», то следует спрятать частицу «бы». Если *que* отсутствует, то оставить по предварительным данным. «Прошедшее время несовершенного вида + бы» дается в таблицах в качестве соответствия для окончаний *subjonctif* в тех группах, где оно отличается от *présent ind.* После правила 4 переходим к правилу 7, определив предварительно род глагола.

Род глагола в прошедшем времени таков же, как у ближайшего слева существительного или местоимения в именительном падеже, с которым французский глагол согласуется в числе и лице, если между ними нет другого глагола; если имеется еще глагол, то следует искать существительное или местоимение в им. падеже дальше влево.

П р а в и л о 5. Случай, когда по предварительным данным получено *participe*.

Причастия (*participe présent* и *participe passé*) могут играть в предложении двоякую роль: они или функционируют как форма глагола (образуя сказуемое) и тогда должны обрабатываться вместе с глаголами в первую очередь, или функционируют как прилагательные и тогда должны обрабатываться после обработки существительных. Поэтому работа над причастиями начинается с выяснения их роли в переводимой фразе.

A. Participe passé

1. Если *participe passé* стоит после любой формы глаголов *avoir*, *s'être* и *être* (последний — для глаголов, имеющих специальное указание на то, что они спрягаются с *être*, например, *aller*, *venir*), причем вспомогательный глагол может быть отделен от причастия наречиями и выражением *le premier*, то *participe passé* считается глагольной формой, которая обрабатывается вместе с глаголами в первую очередь и переводится прошедшим временем совершенного вида. Показатель числа определяется по *avoir*, *s'être* или *être*, сами же формы вспомогательных глаголов при переводе отбрасываются.

2. Если отсутствуют вышеизложенные условия, *participe passé* считается прилагательным (Ap^1) и обрабатывается после существительных.

B. Participe présent

1. Если *participe présent* стоит: а) в начале фразы; б) после *en*; в) после подчинительного союза; г) после глагола (при этом может отделяться наречиями и местоимениями), то *participe présent* считается глагольной формой, обрабатывается вместе с глаголами и переводится деепричастиями настоящего времени.

II. Если отсутствуют указанные условия, *participe présent* обрабатывается после существительных.

а) *participe présent* после существительного в именительном падеже.

1. При наличии справа от *participe* глагола в 3-м лице, согласующегося в числе с этим существительным, от которого он не отделен другим существительным или местоимением в том же числе и в именительном падеже, переводится действительным причастием настоящего времени именительного падежа. В роде и числе это причастие согласуется с предшествующим ему существительным.

2. При отсутствии вышеуказанных условий (нет глагола, или он отделен другим существительным, местоимением) *participe présent* переводится личной формой глагола в 3-м лице. В числе этот глагол согласуется с указанным существительным в именительном падеже, во времени — с ближайшим слева или справа глаголом. II в том случае, когда справа есть еще глагол, перед существительным в переводе ставится союз «когда», если же справа другого глагола нет, то ставится союз «причем», т. е. перевод осуществляется по схеме:

$$S_N + \text{part. prés.} + (\text{subst. adj.}) + \begin{matrix} S_N \\ \text{Pr}_N \end{matrix} + \begin{matrix} \text{наст. буд. пр.} \\ \text{verbe} \end{matrix} =$$

когда наст. буд. пр. сущ. наст. буд. пр.

$$\text{причем} + \text{сущ.} + \text{глагол} + (\text{сущ. прил.}) + \text{местом.} + \text{глагол}$$

В пункте II a1 речь идет о *participe présent*, которое является определением к существительному, а в пункте II a2 — об абсолютном причастном обороте типа: *vous*

¹ Ap — условное обозначение «прилагательных-причастий» в отличие от A — «настоящих» прилагательных.

voyons que l'ordre de $F(z)$ étant ρ' , celui de $h^{(z)}$ ne peut dépasser ρ' (стр. 76). «Мы видим, что, когда порядок $F'(z)$ есть ρ' , порядок $h^{(z)}$ не может превосходить ρ' ».

б) Participe présent после существительного в косвенном надеже.

1. Если participe présent имеет окончание *-ant*, следует руководствоваться правилами 3, 4.

2. Если participe présent имеет окончания: *-ante* (ж. род, ед. число), *-ants* (м. род, мн. число), *-antes* (ж. род, мн. число), то выполнять правила 3 и 4 следует только тогда, когда этому не мешает согласование; в противном же случае относить к ближайшему существительному, которое согласуется с *participe présent* в роде и числе (в третьем случае следует искать его влво, в четвертом — выбрать *substantif* после *de*).

3. Если найденное существительное не является существительным с *de* без детерминатива, *participe présent* переводится, как в пункте 11 а1.

4. *Participe présent* в позиции после *substantif + de + substantif* относится к первому *substantif* и переводится, как в пункте 11 а1.

Participe présent после сочинительного союза переводится так же, как ближайшее слева *participe présent*; если таковое отсутствует, то переводится, как ближайшее слева *participe passé* или прилагательное. После правила 5 переходим к правилу 7.

Правило 6. Неопределенная форма совершенного вида.

Помета неопределенной формы совершенного вида ставится при окончаниях *infinitif*.

а) Если перед французским *infinitif* имеется предлог *sans*, который может быть отделен местоимениями и наречиями, то неопределенную форму следует заменить на «не + деепричастие несовершенного вида».

б) Если перед французским *infinitif* имеется одна из форм глагола *venir*, которая может быть отделена местоимениями и наречиями (имеется в виду *passé immédiat*), то неопределенную форму следует заменить личной — прошедшим временем совершенного вида в том числе, какое имеет *venir*. Перед личной формой ставится «только что», форма же *venir* опускается.

в) Если перед французским глаголом стоит одна из форм глагола *aller*, которая может быть отделена местоимениями и наречиями (имеется в виду *futur immédiat*), то неопределенную форму следует заменить личной — будущим временем совершенного вида, число и лицо такое, как у *aller*; форма *aller* опускается. После правила 6 переходим к правилу 7.

Правило 7. Определение возвратности. Глагол считается возвратным, если перед ним имеются местоимения *je me*, или *nous nous*, или *se* (между местоимениями может стоять отрицательная частица *ne*: *je ne me*, *nous ne nous*), или если в 1-м лице мн. числа местоимение *nous* находится только справа от глагола. После правила 7 переходим к правилу 8.

Правило 8. Определение наличия отрицания.

а) Если перед французским глаголом (за исключением *pouvoir*) имеется *ne* (которое может отделяться посредством *pas* от глагола в *infinitif*), а во фразе имеются *pas*, *point*, *guère*, *rien*, *jamaïs*, *personne*, *aucun*, *plus* (при их поиске не переходить через другое *ne* или *que* — частицу), то глагол считается отрицательным. Если эти условия отсутствуют, глагол считается утвердительным.

б) Если перед глаголом *pouvoir* имеется *ne*, а вправо от глагола нет частицы *que*, то он считается отрицательным. Если эти условия отсутствуют, он считается утвердительным.

«Синтез»

Под «обработкой слова синтезирующими правилами» понимается образование нужных форм русских слов и размещение их во фразе в соответствии с данными, которые получаются при помощи анализирующих правил.

В рабочем участке «памяти» машины перед началом синтеза имеются:

1. Переводящие (русские) основы для всех слов французской фразы, по одной или по несколько для каждого.

2. Указания, в какой форме должно быть то или иное переводящее слово и на какое место следует его поставить. Отсутствие этого последнего указания означает, что слово надо оставить на том же месте, которое занимало соответствующее французское слово.

3. Русская информация к переводящим основам (см. стр. 114).

4. Таблицы русских окончаний.

5. Группа синтезирующих правил.

Синтезирующие правила сравнительно с анализирующими правилами значительно меньше по объему и проще по содержанию. Работа синтезирующих правил сводится к двум операциям:

1. По «указанию о выборе основы», имеющемуся в русской информации, соответствующее правило устанавливает, какую из нескольких основ, переводящих одно

французское слово, следует выбрать для образования нужной нам формы. Например, для перевода слова *permutation* даны две русские основы: 1) «перестановок-» и 2) «перестановок-», а при них «указание о выборе основы» № 2. «Синтезирующее правило» № 2 гласит, что для слов, имеющих такое указание, вторая основа берется для родительного падежа множественного числа, а первая — для всех остальных форм.

Для глагола *considérer* даются основы: 1) «рассматрива-», 2) «рассмотр-» и «указание о выборе основы» № 5. «Синтезирующее правило» № 5 гласит, что для слов, имеющих такое указание, первая основа используется для образования несовершенного вида, а вторая — для образования совершенного вида.

2. По «указанию о выборе окончания» машина обращается к соответствующей таблице окончаний и, согласно данным анализирующих правил, выбирает нужное окончание и присоединяет его к основе. Для причастий выбирается суффикс, а окончание берется из таблиц падежных окончаний прилагательных. Полученные слова расставляются по указанию о порядке слов. Образовавшаяся фраза печатается печатающим устройством.

Так в общих чертах будет осуществляться машинный перевод с французского языка на русский.

Для простоты изложения здесь были опущены многочисленные технические подробности, но об одной из них, довольно существенной, следует упомянуть: это использование так называемых «особенностей». Все слова, имеющие какие бы то ни было индивидуальные особенности в их употреблении, в выборе их перевода, в их влиянии на другие слова и т. д., пронумерованы. Порядковый номер слова и называется его «номером особенности».

Например, «номер особенности» имеет глагол *aller*, поскольку, когда он стоит перед infinitif, то он не переводится, infinitif же переводится личной формой (см. выше стр. 120). Союз *si* также имеет «номер особенности», так как он требует замены настоящего времени будущим (см. стр. 118, правило 16), а будущего времени — прошедшим несовершенного вида + «бы» (см. стр. 118, правило 26). Слова, не имеющие никаких особенностей, имеют нулевой «номер особенности».

Подобная кодировка позволяет формулировать и проверять правила, касающиеся частных особенностей слов, в применении к номерам особенности, а не к самим словам. Таким образом, мы получаем возможность классифицировать и объединять особенности, т. е. если несколько слов имеют совершенно тождественные особенности, то им приписывается общий «номер особенности». Это значительно упрощает все проверки. Так, например, в «Правилах обработки существительных» есть такое правило: «существительное после глаголов *s'appeler, se nommer, devenir, paraître* ставится в творительном падеже». Если не пользоваться «номером особенности», машине придется сделать четыре проверки — сравнить стоящий перед существительным глагол с четырьмя вышеуказанными, чтобы установить, не является ли он одним из них. Мы же приписываем всем им общий «номер особенности», например 157, и формулируем правило для машины так: «Существительное после глагола с № 157 ставится в творительном падеже». Теперь машина должна делать только одну проверку — не имеет ли предшествующий глагол номера 157.

*

Вышеизложенные правила сформулированы на основе французских математических текстов¹.

Настоящая работа является первым опытом машинного перевода с французского языка на русский и не претендует на полноту и окончательность результатов.

О. С. Кулагина и И. А. Мельчук

ГРАНИЦЫ ПРИМЕНИМОСТИ МАШИННОГО ПЕРЕВОДА

Машинный перевод, т. е. возможность перевести посредством машины текст с одного языка на другой, в настоящее время является реальностью. Разумеется, это не значит, что человека-переводчика можно или нужно заменить машиной. Машинный перевод будет целесообразен только в известной ограниченной области переводческой работы, и языковеды должны точно уяснить себе, в какой именно области и для

¹ P. Appel, *Traité de mécanique rationnelle*, tt. I, II, IV, Paris, 1926 — 1937; E. Borel, указ. соч.; егo же, *Leçons sur les fonctions entières*, Paris, 1921; H. Lebesgue, *Leçons sur l'intégration et la recherche des fonctions primitives*, Paris, 1928; E. Picard, *Traité d'Analyse*, tt. I — III, Paris, 1922 — 1928.