

ПРИКЛАДНОЕ ЯЗЫКОЗНАНИЕ

ИССЛЕДОВАТЕЛЬСКАЯ РАБОТА В ОБЛАСТИ МАШИННОГО ПЕРЕВОДА В КИТАЙСКОЙ НАРОДНОЙ РЕСПУБЛИКЕ¹

Проблемам машинного перевода отдельные китайские научные работники стали уделять внимание только с 1956 г.² Более широкий интерес к проблемам МП возник в Китае после того, как китайские специалисты ознакомились с результатами экспериментального перевода с английского на русский, осуществленного в СССР на БЭСМ.

Исследования в области МП начались собственно с 1958 г.: в Институте вычислительной техники АН КНР была создана группа по изучению МП, которая в тесном сотрудничестве с организованной в том же году специальной группой Института языкознания АН КНР занялась разработкой алгоритма МП с русского языка на китайский; одновременно указанная группа Института языкознания КНР в контакте с Институтом иностранных языков, где в марте 1959 г. была создана группа по англо-китайскому МП, изучает проблемы машинного англо-китайского перевода. В декабре 1958 г. аналогичные группы возникли в Институте русского языка и в Политехническом институте Южного Китая в Гуанчжоу. Кроме того, разработкой проблем машинного перевода занимаются и предполагают заниматься Институт научной информации АН КНР, Харбинский политехнический институт и ряд других институтов.

В настоящее время основное внимание уделяется составлению русско-китайского и англо-китайского алгоритмов. Работа по МП с французского и немецкого языков находится в стадии подготовки. Сравнительно ощутимые результаты достигнуты по составлению русско-китайского алгоритма. Значительная работа проведена по подготовке англо-китайского машинного перевода, разработка алгоритма которого осуществляется методом синтагматического анализа. В текущем году предполагается осуществить опытный перевод на машине.

В Китайской Народной Республике осуществляется подготовка кадров для работы в области МП, в частности были проведены научные семинары на темы: 1) «Машинный перевод как одна из отраслей прикладного языкознания», 2) «Основные принципы устройства и работы электронной вычислительной машины и основы программирования», 3) «Методика составления русско-китайского словаря для МП», 4) «Грамматический анализ в системе алгоритма бинарного машинного русско-китайского перевода», 5) «Грамматический синтез в системе алгоритма бинарного машинного русско-китайского перевода», 6) «Проблема языка-посредника». Читался также специальный курс «Программирование алгоритмов машинного перевода», который был рассчитан не только на математиков, но имел своей целью также ознакомить лингвистов с основами программирования. Некоторые результаты исследовательской работы в области МП были представлены на выставке ноября 1958 г., которая была организована высшими учебными заведениями.

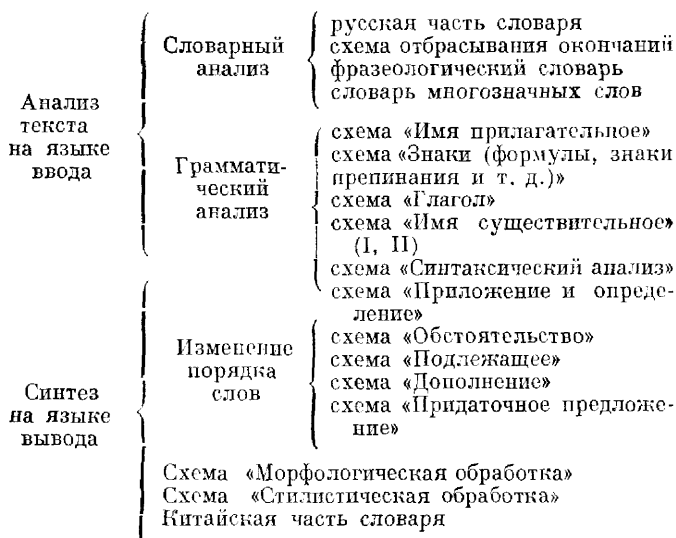
Исследования в области МП начались в Китае с разработок проблематики машинного русско-китайского перевода на базе математической литературы. В качестве объекта исследования были взяты книги А. Я. Хинчина «Краткий курс математического анализа» (при этом основной упор делался на изучение проблемы изменения порядка слов при переводе) и И. Г. Петровского «Лекции об упрощениях с частными производными». В основу исследования был положен метод морфологическо-структурного анализа (а не метод синтагматического анализа, как в англо-китайском алгоритме).

В течение года составлен первоначальный вариант алгоритма машинного русско-китайского перевода, включающий в себя: 1) словарь (2045 однозначных слов, 152 фразеологических единицы, 206 многозначных слов); 2) 6 схем анализа; 3) 7 схем синтеза. В настоящее время мы занимаемся уточнением и усовершенствованием этого варианта алгоритма и подготовкой его к программированию. Ниже приводятся составные элементы русско-китайского алгоритма³, расположенные в рабочей последовательности.

¹ Сообщение, сделанное на Советании по математической лингвистике (Ленинград, 15—21 апреля 1959 г.).

² См.: Ю й Да о - ц ю а н ь, Механизация перевода «Вестник Центр. Академии нац. меньшинств», 1956, сентябрь (на китайск. яз.); Д ю Ю н-ц ю а н ь, О машинном переводе «Сообщения о лингвистических исследованиях», № 9, 1957 (на китайск. яз.).

³ В настоящее время в этот алгоритм уже внесен целый ряд уточнений и изменений.



В ходе практической работы по составлению русско-китайского алгоритма мы тщательно сопоставляли лексические и грамматические явления русского и китайского языков. В частности, детальному изучению подверглись следующие проблемы:

1. С л о в а р н а я ф о р м а. Разрабатывались два варианта словаря: словарь основ и словарь слов. Существенным недостатком словаря основ является обилие омографов. В словаре основ омографами выступают не только такие пары слов, как: «перевод» — «переводить», «так» — «так-ой», но и пары слов типа «работ-а» — «работ-ать», «отдельн-о» — «отдельн-ое». Словарный анализ при помощи словаря слов осуществляется медленнее, но в этом случае почти полностью исключаются омографы. В конечном итоге мы стали включать в словарь машинного перевода слова в словарной форме, но в ряде случаев не отказывались и от использования здесь основ.

2. З а м е н а о к о н ч а н и й с о о т в е т с т в у ю щ е й и н ф о р м а ц и - е й. После словарного анализа все окончания заменяются окончательной или альтернативной информацией. Так, окончание *-ли* заменяется окончательной информацией (прошедшее время, множественное число, сказуемое), окончание *-у* — альтернативной информацией (дат. падеж, ед. число, или вин. падеж, ед. число, или предл. падеж, ед. число). Отдельные окончания (*-ся*, *-о*, *-ее*) заменяются специализированной информацией. Этот метод позволяет экономить ячейки памяти и облегчает проведение анализа.

3. П о р я д о к с л о в. Практика составления и проверки алгоритма подтвердила пять основных принципов, которые были положены в основу работы схемы «Изменение порядка слов». Эти пять принципов сводятся к следующему: а) при изменении порядка слов в первую очередь необходимо установить основную ось изменения порядка слов — ею, по нашему мнению, является сказуемое; позиция подлежащего, дополнений и части обстоятельств находится в зависимости от местоположения сказуемого, в определительном словосочетании такой осью является определяемое; б) чрезвычайно важна последовательность работы схем алгоритма и последовательность операций внутри каждой отдельной схемы, ибо неправильная последовательность здесь приводит к многочисленным ошибкам при переводе; в) определение и определяемое, равно как и другие взаимосвязанные синтаксические элементы, необходимо рассматривать как единое целое (единый комплекс) и снабжать единым кодовым числом, в противном случае неизбежны искажения перевода или полная утрата смысла; г) знаки препинания фиксируются или снимаются в ходе работы схемы «Изменение порядка слов»; д) морфологические элементы языка ввода (в данном случае русского языка) в известной своей части передаются посредством введения китайских служебных или знаменательных слов в процессе изменения порядка слов.

4. О б о р о т ы с п р е д л о г о м. В русском языке предлогные обороты синтаксически многозначны, причем в ряде случаев различение их синтаксической многозначности представляет значительные трудности. Синтаксические функции предлогов определяются нами одновременно с анализом их полисемии. Так, например, перевод и синтаксические функции предлога *в* после обработки оказываются определенными следующим образом: а) *цзай... ли*, вводит обстоятельство¹; б) «не переводится», вводит косвенное дополнение; в) «не переводится», вводит определение; г) *вай*, вводит косвенное дополнение; д) *цзай*, вводит обстоятельство; е) *дао*, вводит косвенное дополнение ж) *цзай... шан*, вводит обстоятельство, и т. д.

¹ Здесь дается китайский перевод русск. *в* и указывается его синтаксическая функция.

5. С л о в а р ь ф р а з е о л о г и ч е с к и х о б о р о т о в. Фразеологические обороты состоят из двух, трех или четырех слов. Лексические единицы, входящие в состав того или иного фразеологического оборота, обладают морфологической инвариантностью, что удобно для анализа. Фразеологические обороты подразделяются на два типа: 1) независимые, т. е. те, которые не управляют и не управляемы другими словами фразы (*так как, в такой случай*), 2) зависимые, которые в свою очередь разделяются на два подтипа: а) управляющие (*в отличие от*), б) управляемые (*уравнение с частный производная*). Фразеологическим оборотам в зависимости от принадлежности к тому или иному типу присваиваются соответствующие признаки. Например, фразеологический оборот первого типа *так как* снабжается признаком «союз неоднородный»; фразеологические обороты второго типа получают признак в зависимости от характера управления: фразеологический оборот *в отношении* снабжается признаком «предлог, управляет родительным падежом, вводит обстоятельство», фразеологический оборот *уравнение с частный производная* — признаком «оставить первое слово с приданной ему информацией».

Кроме перечисленных, из проблем, которые частично уже исследованы, или находятся в стадии изучения, или намечены к разработке, назовем следующие: а) структурно-семантический анализ предлогов; б) перевод определительных придаточных предложений, вводимых союзным словом *который*; в) перевод морфологических элементов (перевод глаголов в личной форме, перевод причастий, деепричастий и определение их позиции в предложении; перевод имен существительных в родительном падеже в функции определения и т. д.); г) анализ эллиптических предложений; д) анализ местоимений; е) передача числа имен существительных; ж) перевод числительных (проблема сочетания числительных с той или иной категорией счетных слов; з) закономерности ввода новых или усечения имеющихся слов при переводе; и) анализ запятых; к) исследование обобщенного значения слова; л) исследование закономерностей стилистики и т. д.

Чрезвычайно важное значение в разработке проблем МП имеет статистическое исследование. Проведено статистическое исследование частотности окончаний и слов в русском языке. Справедливости ради следует отметить, что статистические исследования занимают до последнего времени незаслуженно малое место в нашей работе над машинным переводом. Так, в частности, совершенно не проводились статистические исследования частотности употребления грамматических категорий.

Результаты проведенных исследований были обобщены в ряде статей и лекций¹. В заключение нужно отметить, что успешное развертывание в нашей стране исследовательской работы в области МП неотделимо от помощи советских специалистов, передающих нам свой опыт и в этой новой отрасли науки.

Л ю Ю н - ц ю а н ь

¹ Перечислим важнейшие из них: Л ю Ю н - ц ю а н ь, Вопрос о порядке слов и его решение при машинном переводе с русского языка на китайский, журн. «Юйняньцзю» («Лингвистические исследования»), 1959, № 4; Л и М и ц, П э н Ч у - ч у н ь, Применение вычислительных машин для автоматического перевода, журн. «Дяньцзю цзисуаньцзю дунтай» («Новости электронной вычислительной машины»), 1959, № 1; Л ю Ю н - ц ю а н ь, Машинный перевод, журн. «Чжунго юйвань», 1958, № 12 и т. д.