
О РОЛИ А. А. ЛЯПУНОВА В РАЗВИТИИ РАБОТ ПО МАШИННОМУ ПЕРЕВОДУ В СССР

О. С. КУЛАГИНА

(МОСКВА)

I

Машинный перевод (МП) и математическая лингвистика принадлежат к числу научных областей, в становлении которых Алексей Андреевич Ляпунов активно участвовал и на развитие которых оказывал большое влияние на протяжении многих лет.

Работы по МП в СССР начались в 1954 г., когда ЭВМ в нашей стране насчитывались единицами, их технические возможности были еще очень малы, возможные сферы применения неясны. В этот период предложение использовать ЭВМ для решения такой сложной невычислительной задачи, как перевод текстов с одного языка на другой, многим казался неосуществимым. Большинство лингвистов первоначально отнеслись к идее МП с недоверием. Алексей Андреевич очень много сделал для преодоления этого недоверия и налаживания работ. Это было время, когда кибернетика в нашей стране еще только зарождалась, причем ее первые шаги часто встречали сопротивление. В этот период личные качества Алексея Андреевича, его энергия, умение увлекать людей новыми задачами и перспективами, ораторский талант сыграли как для кибернетики в целом, так и для МП, в частности, огромную положительную роль.

Для того чтобы начались работы по МП, Алексей Андреевич выступал с докладами в Институте языкознания АН СССР и МГПИИЯ, организовывал совещания лингвистов и математиков, проводил беседы с людьми разных специальностей. Ему удалось заинтересовать этой новой тематикой ряд лингвистов старшего поколения и привлечь к непосредственной работе молодых лингвистов. В Институте прикладной математики АН СССР, где А. А. Ляпунов работал в то время, он создал группу из своих учеников-математиков, которые начали работу в области МП сначала с французского языка на русский, а затем и с английского языка на русский.

Однако Алексей Андреевич не ограничился созданием группы в том институте, где он сам работал. Его активная деятельность явилась тем мощным импульсом, в результате которого довольно быстро одна за другой в разных городах нашей страны возникли группы, занимающиеся МП. Можно без преувеличения сказать, что своим возникновением в нашей стране машинный перевод обязан А. А. Ляпунову.

В первые месяцы работы группы, созданной А. А. Ляпуновым в ИПМ, над МП с французского языка на русский для математических текстов Алексей Андреевич интересовался всеми подробностями буквально ежедневно; сам он всегда с радостью был готов использовать свое прекрасное

знание французского языка для решения любых вопросов вплоть до самых мелких, как например, приписыванию некоторому слову каких-то признаков или выбор для какого-то слова перевода, наиболее пригодного в текстах по математике.

Впоследствии, когда работы развернулись и группа приобрела опыт, Алексей Андреевич отошел от решения мелких вопросов, однако все то время, пока А. А. Ляпунов жил в Москве, эта группа работала под его непосредственным руководством, а после отъезда Алексея Андреевича в 1962 г. в Новосибирск — в самом тесном контакте с ним.

В своих самых первых беседах с учениками о машинном переводе А. А. Ляпунов говорил о переводе путем извлечения некоторым способом смысла текста и его передачи на другом языке. Однако сразу же стало ясно, что тогда, т. е. двадцать лет назад, реализовать построение алгоритма, ведущего перевод через построение семантической структуры текста (как это принято говорить сейчас), было невозможно. Слишком много трудностей и неясностей было в том, как надо строить формализованные алгоритмы для работы с текстами, какие словари надо вводить в машину и с какой информацией, какого рода представления информации о переводимом тексте нужно строить, какие лингвистические закономерности можно при этом использовать и каковы вообще эти закономерности и т. д. Выяснилось, что лингвистика совершенно не располагает ни фактическим материалом, ни идеями и представлениями, нужными для построения алгоритмов, которые в каком бы то ни было виде использовали бы смысловое представление текста. В неразработанном состоянии оказалась не только семантика, но и синтаксис: так, ни для одного языка не существовало перечня синтаксических конструкций, не были изучены условия их сочетаемости и взаимозаменяемости, не были разработаны правила образования более крупных структурных единиц из более мелких и т. д. Вообще классическая лингвистика оказалась несостоятельной перед лицом новых задач, вставших в связи с началом исследований по машинной переработке текстов на естественных языках. В сущности, ни на один вопрос, поставленный в связи с построением алгоритмов переработки текстов, нельзя было получить точного ответа. Даже фактических данных о тех или иных явлениях языка в лингвистике либо совсем не было, либо имеющийся материал был изложен так бессистемно, на таком уровне формально-логической мысли, что извлечь фактические данные из лингвистических работ, как правило, оказывалось просто невозможным.

Решить сразу все возникшие проблемы было немислимо, поэтому естественно, что группа МП в ИПМ, руководимая А. А. Ляпуновым, пошла по пути постепенного преодоления трудностей. Первый алгоритм МП, который начал работать в 1956 г., — алгоритм перевода математических текстов с французского языка на русский ФР-I, — строился эмпирически и был, конечно, весьма не совершенным. Однако он дал материал для дальнейшего развития работ в различных направлениях. Следующий алгоритм, который строился в ИПМ также под руководством А. А. Ляпунова, — алгоритм МП с английского языка на русский, уже имел более четкую логическую структуру (см. [15]).

В дальнейшем, как уже было сказано, и после отъезда Алексея Андреевича из Москвы, общее руководство работами по МП в ИПМ оставалось за ним.

Нужно сказать, что интерес А. А. Ляпунова к МП никогда не ограничивался теми исследованиями, которые велись под его непосредственным руководством. В его поле зрения находилась также деятельность других коллективов, которые со второй половины 50-х годов начали работать над проблемами МП в Ленинграде, Киеве, Ереване, Тбилиси и других горо-

дах. Как и во многих других областях кибернетики, Алексея Андреевича занимало в первую очередь общее направление хода дел, общая стратегия в данной области, а также ее связи как с другими областями кибернетики, так и с лингвистикой. Эту общую линию А. А. Ляпунов ясно представлял и видел далеко вперед.

Нужно сказать, что проведение такой общей линии было делом далеко не простым, поскольку МП, как всякая новая область, возникшая на стыке сложившихся ранее научных областей, испытывал на себе разнообразные, нередко весьма противоречивые воздействия. Трудности возникали как со стороны лингвистов, так и со стороны математиков.

Что касается лингвистов, то даже у тех из них, кто хотел работать по-новому, и готовы были разрабатывать лингвистические проблемы в нужном МП аспекте, не всегда был достаточно четко и правильно налаженный контакт с математиками, что приводило к отсутствию у лингвистов ясного понимания процессов и алгоритмов, для которых они должны были дать лингвистический материал. Отсюда возникла неясность в том, какие именно лингвистические данные и под каким углом зрения им надо собрать и в какую общую схему их следует уложить. В свою очередь отсутствие общей картины и неразработанность слишком большого числа проблем приводили к тому, что иногда лингвисты начинали увлекаться решением частных вопросов, отходя на позиции, близкие к традиционно лингвистическим, забывая о целях МП и еще больше ослабляя контакт с математиками.

Что касается математиков, то и здесь не все обстояло просто, поскольку математики столкнулись с совершенно новым, чрезвычайно сложно организованным, трудно поддающимся осмыслению и формализации объектом. Работа с ним требовала создания новых математических подходов и методов, тогда как у некоторых математиков возникала тенденция перенести в эту новую сферу какой-либо привычный им математический аппарат, причем без достаточно глубокого проникновения в суть и специфику лингвистических феноменов.

Кроме вопроса взаимодействия математики и лингвистики на почве МП, другим пунктом противоречий было соотношение теоретического и практического подходов к МП. Здесь были свои крайности. С одной стороны — чрезмерное увлечение теорией, при котором МП рассматривался только как источник и стимул теоретических исследований сугубо академического стиля. С другой стороны преждевременный практицизм, т. е. необоснованное стремление к поспешной реализации систем перевода, представляющих собой перечисление частных случаев и ситуаций, не объединенных общими идеями и понятиями, где собственно алгоритм сводился к перебору этих частных случаев.

Именно в силу наличия этих противоречивых тенденций и увлечений так велика роль А. А. Ляпунова в становлении МП в СССР, поскольку он, как никто другой, заботился о том, чтобы МП развивался как единое, цельное, правильно сбалансированное направление. О таком едином МП Алексей Андреевич писал в своих печатных работах (здесь выделяются работы [12] и [13]) и говорил в устных выступлениях.

МП для А. А. Ляпунова был прежде всего составной частью кибернетики. Некоторые из его высказываний о необходимости тесной связи МП с кибернетикой и использовании подходов, развитых в других областях кибернетики, для нахождения рациональной структуры алгоритмов МП приводятся ниже.

Алексей Андреевич постоянно подчеркивал необходимость тесного сотрудничества математиков и лингвистов на почве МП при ведущей роли математиков, причем именно математиков-кибернетиков, т. е. связанных

с ЭВМ и интересующихся созданием нового математического аппарата для сложных систем переработки информации.

Из сказанного выше должно быть понятно, почему А. А. Ляпунов критиковал те работы, в которых определяющей была не математико-кибернетическая, а лингвистическая сторона, т. е. нахождение рациональной структуры переводческих алгоритмов и изучение закономерностей преобразования информации занимали подчиненное положение по отношению к лингвистическому аспекту задачи. Ясно должно быть также, почему он не ждал большого вклада в МП от тех работ, где к лингвистике применялся аппарат математической логики и теории алгоритмов с их интересом к алгоритмической разрешимости или неразрешимости задач, сформулированных для бесконечных множеств, в то время как трудности нахождения эффективных алгоритмов МП, связанные с ростом переборных, не пропадают при отказе от бесконечности рассматриваемых множеств.

А. А. Ляпунов уделял много внимания правильному соотношению теоретических исследований в МП и его практической реализации. Некоторые из его высказываний на эту тему приведены ниже.

Количество опубликованных работ А. А. Ляпунова, относящихся к МП и математической лингвистике, не очень велико. Несколько статей было написано им в начальный период становления МП в СССР — 1955—1957 годах (см. [1], [2], [4]), пять работ — в период с 1958 г. по 1970 г. ([5], [7], [8], [10], [11]). Разделы, посвященные МП, есть также в публикациях [3], [6], [9]. За исключением обзорной статьи в сборнике «Математика в СССР за 40 лет» [6], все это — публикации с соавторами, чаще всего с учениками. В 1971 г. вышли две работы: «О некоторых методологических вопросах, относящихся к машинному переводу» [12] и «О задачах математической лингвистики» [13], — идеи и установки которых принадлежат А. А. Ляпунову, хотя они написаны совместно с учениками. К сожалению, обе они вошли в сборник, который вышел тиражом всего 300 экземпляров. В 1972 г. А. А. Ляпунов написал послесловие к статье автора в сборнике «Проблемы кибернетики» [14].

Но надо сказать, что воздействие А. А. Ляпунова на машинный перевод лишь в незначительной степени определялось его публикациями. Впрочем, это относится не только к МП, но и к кибернетике в целом. Непосредственное воздействие Алексея Андреевича на всех окружающих было неизмеримо сильнее того влияния, которое оказывали его напечатанные работы. Алексей Андреевич очень много времени проводил в активном общении с коллегами, учениками, с самыми разными людьми. Так и в области МП свои идеи, соображения, замыслы А. А. Ляпунов излагал в докладах на конференциях, лекциях, выступлениях на совещаниях, беседах с учениками, во всевозможных обсуждениях, разговорах, отзывах, письмах или в виде записок, которые он часто начинал диктовать во время разговора, а иногда писал сам.

Ниже приводится ряд выдержек из таких материалов. Хочется обратить внимание читателя на следующее важное обстоятельство. Приводимые отрывки не готовились Алексеем Андреевичем для печати, не редактировались, не шлифовались им. Читателю надо не упускать из виду, что это именно сырой, необработанный, не доведенный до окончательной редакции материал. Поэтому, с одной стороны, в приводимых материалах именно в силу обстоятельств их возникновения некоторые мысли выражены не всегда с исчерпывающей полнотой. Нередко в одном случае, в зависимости от контекста, Алексей Андреевич делал акцент на одной стороне вопроса, а в другом случае, говоря о том же самом, сосредоточивал внимание на другой стороне. Подробные формулировки следует воспринимать

как дополняющие друг друга. С другой стороны, в приводимых отрывках есть довольно много повторений, нередко эти выдержки перекликаются с тем, что вошло в две упомянутые работы 1971 г. [12], [13]. Однако мы не убрали этих повторений, поскольку, во-первых, хотели сохранить самый стиль Алексея Андреевича, а, во-вторых, как нам кажется, нередко те же мысли, которые изложены в статьях [12, 13], сформулированы в приводимых отрывках в более сжатой и вместе с тем более решительной и яркой форме. Для читателя, который хочет иметь достаточно полное представление о точках зрения А. А. Ляпунова на МП и математическую лингвистику, самое лучшее — это внимательно сопоставить приводимые ниже отрывки с опубликованными работами А. А. Ляпунова, в первую очередь со статьями [12] и [13].

Для данной статьи были отобраны те места из сохранившихся записок, отзывов, писем и т. д., где говорится о проблематике МП, методологии МП, о связях МП с кибернетикой и лингвистикой. Такой выбор продиктован двумя соображениями. С одной стороны, эти высказывания А. А. Ляпунова еще долгое время будут оставаться актуальными. С другой стороны, именно эти вопросы определяют «генеральную линию» и перспективы МП как научного направления, а нам кажется, что именно очерчивание МП как научного направления и определение перспектив его развития составляет основной вклад А. А. Ляпунова в развитие МП в нашей стране.

II

О проблематике машинного перевода как научного направления, о его связях с другими областями науки Алексей Андреевич говорил и писал неоднократно, как и в статьях об МП (см., например, [12]), так и в статьях или докладах на общекибернетические темы (см., например, [19]). Ниже приводятся строки записки, написанной Алексеем Андреевичем еще в 1968 г., в которой эти проблемы сформулированы кратко, но, вместе с тем, настолько широко, что воплощение в жизнь очерченной здесь программы потребует еще многих лет напряженной исследовательской работы. Эта записка состоит из двух разделов.

Некоторые общие соображения по поводу МП

1. МП является органической составной частью кибернетики. Главная направленность: практическая реализация МП и выработка эффективных способов работы реально существующих ЭВМ с человеческим языком, а также выработка новых требований к ЭВМ с тем, чтобы возможности использования человеческих языков могли быть существенно улучшены.

Такая постановка задачи имеет большое экономическое, общекультурное и научное значение.

2. Как и все разделы прикладной кибернетики, МП имеет научные связи двоякого типа: внутрикибернетические — строение алгоритмов, приближенная алгоритмизация, машинная реализация, машинный эксперимент, вопросы формализации и кодировки; лингвистические — выработка точной лингвистики: принципы классификации слов и отношений между словами в пределах текста, изучение иерархического членения текста на относительно автономные элементы, общие принципы объединения этих элементов, по возможности сопоставление структурных особенностей разных уровней. Крайняя необходимость установления по возможности объективных критериев для оценки лингвистических концепций.

3. Необходимо считаться с тем, что выяснено в смежных областях кибернетики.

Возможность построения иерархии все более полных приближенных алгоритмов.

Необходимость разумного приспособления к особенностям ЭВМ.

Нецелесообразность буквального копирования действий человека.

Необходимость учета данных, полученных при изучении модельных объектов.

Возможность исследования компромиссов между трудоемкостью и качеством алгоритмов.

К вопросу о стратегии в области машинного перевода

1. Работы нужно вести на двух основных рубежах:

- а) Перспективные научные исследования.
- б) Реализация цельного алгоритма.

Теоретические исследования, не подкрепленные реализацией, имеют тенденцию отрыва от реальных задач и обращения в новый вариант классической лингвистики. Именно большой эксперимент, направленный на реализацию, может давать реальные постановки теоретических задач. Отсутствие перспективы реализации компрометирует направление и блокирует его основные научные связи с кибернетикой.

2. Теоретические исследования должны вестись в двух направлениях:

а) Изучение структуры переводческих алгоритмов: разработка методов оценки их эффективности и трудоемкости, выяснение роли приближенных алгоритмов, привязка к идеям теории локальных алгоритмов и сопоставление с другими алгоритмами, в которых осуществляется «перевод» одного класса объектов в другой класс объектов, например, синтез контактных схем, реализующих данную функцию алгебры логики, упрощение д. н. ф., трансляторы в программировании.

б) Развитие точной лингвистики с неременным экспериментальным опробованием общих концепций и результатов. Ограничение универсальности: возможность приближенных подходов.

Как видно из сказанного, МП был для Алексея Андреевича составной частью кибернетики. Место МП в кибернетике Алексей Андреевич очертил следующим образом в одном из писем, написанных в январе 1968 г.

Содержание кибернетики

I. Общие основы или принципы изучения управляющих систем и процессов управления (т. е. систем, перерабатывающих информацию с целью выбора образа действия, см. основную схему управления).

II. Наблюдения процессов управления действительности *).

III. Эксперимент с реальными управляющими системами с целью узнать их функционирование в тех или иных условиях или при том или ином изменении структуры.

IV. Моделирование или эксперименты с моделями — при наличии описания реальной управляющей системы и комплекса гипотез о ее работе — построение другой управляющей системы (аналоговые или цифровые машины) с целью сопоставления модели и действительности в сфере функционирования.

V. Формулировка задач общетеоретической природы с целью улучшения моделей или их приближения к действительности — точные классификации, количественный подход.

VI. Использование полученных результатов в конкретных областях при работе с управляющими системами.

Может быть, целесообразно в каждом из этих пунктов выделить свои подпункты, например, алгоритмизация — в пункте «эксперимент с моделями», а также в пункте VI. Оптимальные решения в п. V. Кодирование информации в п. п. II—IV и его изучение в п. V.

Все эти пункты тесно связаны между собой. Возможны переходы от одних к другим. В конкретных задачах и областях, видимо, полезно проследивать, что относится к каждому из пунктов. Каждый требует развития своих подходов. Интересно горизонтальное сопоставление того, что делается в разных областях. В частности, создание алгоритмов МП относится к разделу VI, а попутно возникают задачи, вероятно, из всех остальных пунктов.

Алексей Андреевич неоднократно подчеркивал необходимость рассмотрения разных разделов кибернетики с единых позиций. Сам он обладал удивительной широтой, мог как никто другой сопоставить различные направления кибернетических исследований, провести параллели, указать аналогии, выделить то общее, что является характерным для кибернетики, и это помогало ему видеть далекие перспективы развития кибернетики. Алексей Андреевич был очень обеспокоен наметившейся в последние годы в кибернетике тенденцией к обособлению различных направлений. Он

*) Имеются в виду процессы, протекающие в действительности.

всегда призывал своих учеников к тому, чтобы, работая в разных разделах кибернетики, они уделяли внимание не только стоящим перед ними конкретным проблемам, но и вопросам общеметодологического характера, вопросам связи разделов кибернетики друг с другом. В одном из писем (январь 1968 г.) Алексей Андреевич, говоря о методологии МП, писал:

Задачи методологии современные и конкретные — сделать прогноз о том, какие подходы к той или иной области науки перспективны. Для этого нужно рассматривать, по возможности, много областей, в которых имеется родственная обстановка. Для МП таковыми могут быть: информационные системы, теория управляющих систем, «большие» системы, кибернетические вопросы техники и биологии.

Придавая большое значение внутрикибернетическим связям, А. А. Ляпунов считал также важным, чтобы правильным образом развивались связи тех или иных разделов кибернетики с близкими им другими областями науки. Для МП такой областью является лингвистика.

А. А. Ляпунов часто повторял, что главным, ведущим, в МП является математико-кибернетический аспект. Он всегда подчеркивал, что, хотя основным препятствием для быстрого и успешного развития работ по МП явилось отсутствие той точной лингвистики, о которой говорилось выше, и таким образом основные трудности в МП, как правило, создает его лингвистическая часть, из этого вовсе не следует, что эта лингвистическая часть является в МП основной, самой существенной. Об этих вопросах, а также о необходимости постоянного тесного контакта между математиками и лингвистами в работах по МП Алексей Андреевич писал в письме в январе 1968 г. следующее.

Если общие основы кибернетики в целом пригодны для МП, то общие основы классической лингвистики оказались в целом непригодны для МП. Нужно подчеркнуть, что основная идейная сторона МП лежит в математико-кибернетической сфере и что лингвистические работы в интересах МП должны вестись в тесном контакте с математиками и что недооценка алгоритмической стороны уводит лингвистические работы в сторону.

Как известно, МП привел к возникновению в лингвистике новых подходов. Более того, каждый новый заход в построении переводческих алгоритмов требовал и требует расширения сведений о языке, привлечения новых данных, а следовательно, проведения обширных лингвистических исследований. Однако, как это уже говорилось, Алексей Андреевич придавал большое значение тому, чтобы эти лингвистические работы не отрывались от почвы МП, чтобы новые лингвистические концепции и разработки включались в системы перевода и проходили бы в них экспериментальную проверку.

Я совсем не отрицаю важности новых работ по лингвистике (помня о том, что они импульсированы МП). Однако я считаю, что эта сторона дела отнюдь не заменяет работ по МП, что одни только работы по новой лингвистике и за 300 лет не приведут к реальному МП, что для дальнейшего развития самого МП, а также для дальнейшего развития работ по лингвистике необходим тесный контакт математиков с лингвистами. Когда лингвисты отрываются от математиков, они уходят в дебри, далекие от реализации.

Вопросы, о которых шла¹ речь выше, затронуты еще в одной записке, которая также была написана в 1968 г. Хотя в ней часто повторяются точки зрения и формулировки, бывшие в отрывках, приведенных выше, мы не делали в ней сокращений, чтобы не нарушать ее цельности.

К вопросу о взаимоотношении математики и лингвистики на почве МП

1. Для получения МП необходимы:

- а) достаточная степень изученности языка (это коренная задача лингвистики)*);
- б) разработка алгоритмов перевода на базе сведений о языке, учета возможностей машин и здравых представлений о структуре алгоритма (типичная задача кибернетики);
- в) разработка рациональных методов усовершенствования алгоритмов на основе машинных экспериментов (опять задача кибернетики);
- г) выяснение новых требований к изучению языка (здесь переплетаются соображения кибернетики и лингвистики, причем ведущими являются первыми);
- д) разработка методов программирования переводческих алгоритмов и реальное построение программ (также задача кибернетики);
- е) проведение машинных экспериментов и анализ результатов (в основном задача кибернетики, целесообразно участие лингвистов);
- ж) составление словарей с необходимой информацией (идейная часть — структура информации — смешанная задача лингвистики и кибернетики, техническая часть — составление словарей — задача лингвистики);
- з) очерчивание характера текстов, на которые рассчитан перевод (задача кибернетиков).

2. Для выполнения такой сложной задачи, очевидно, требуется большой коллектив. Необходимо обеспечить разумное распределение заданий внутри коллектива и расстановку сил. Для этого необходимо видеть всю задачу в целом. Такие обязанности падают всегда на долю кибернетиков.

3. Уже из предыдущего видно, что руководящую роль, определяющую, в работах по МП должны выполнять математики-кибернетики.

4. Фактически в работе по МП было обнаружено, что существующие в лингвистике способы описания языка не могут служить основой разработки алгоритмов перевода ввиду их недостаточной отчетливости, а также ввиду того, что их использование предполагает априорное понимание смысла текста. Ввиду этого пункт 1а) пришлось очень скоро заменить пунктом 1г), и требования к описанию языка пришлось разрабатывать одновременно с разработкой алгоритмов. После первых серий экспериментов стали проявляться новые требования к описанию языка, а также новые интуитивные представления о том, какими эти описания должны быть. Это вызвало к жизни ряд весьма интересных работ лингвистов (Мельчук, Жолковский, Мартемьянов). Эти новые способы описания языка лучше соответствуют требованиям МП, но, по-видимому, это направление далеко не исчерпано. Складывается представление, что описание языка распадается на ярусы и для каждого типа алгоритмов целесообразно ограничиваться некоторыми определенными ярусами описания языка.

5. В то же время эти описания языка отнюдь не заменяют алгоритмов перевода, тогда как у лингвистов нередко складывается превратное впечатление, что разрабатываемые ими описания языка представляют собой готовые алгоритмы или способы их замены.

6. К сожалению, в настоящее время в области МП лингвисты более активны, чем кибернетики, и нередко оказывается, что математики идут на поводу у лингвистов, вместо того, чтобы определять стратегическое направление этой области. Именно это обстоятельство, наряду с весьма неблагоприятной административной бюрократической конъюнктурой, сильно задерживает реализацию МП (сравнение с ходом дел в Гренобльской группе).

7. Общую стратегию в МП должны определять математики, исходя из стремления к реализации МП. При этом реализация должна вестись поэтапно. Начинать нужно с ограниченных текстов — ограничение области, которое ведет, с одной стороны, к большому облегчению решения вопросов об омонимии, сужению семантики, а в некоторой степени и ограничению синтаксических структур. Кроме того, это позволит начать работу с меньшим объемом словаря, а затем нужно постепенно расширить возможности и, в частности, дифференцировать алгоритмы по областям. Не следует бояться того, что некоторая часть фраз будет переводиться неудачно.

8. Весьма существенно то, что если МП заработает хотя бы в ограниченных пределах, то можно рассчитывать на то, что это повлияет на стилистику научных статей. Как только МП окажется в состоянии ускорять распространение информации,

*) К сожалению, как это неоднократно отмечалось, в том числе и в данной статье, реальная ситуация в лингвистике совершенно не соответствует этому требованию. Слишком многие проблемы и явления из области синтаксиса и семантики языков лингвистами раньше вообще не изучались. Поэтому в действительности работы в области создания алгоритмов МП и продвижение в области изучения естественных языков переплетены очень сложным образом.

К вопросу о взаимоотношении математики и лингвистики на почве МП

1. Для получения МП необходимы:

- а) достаточная степень изученности языка (это коренная задача лингвистики) *);
- б) разработка алгоритмов перевода на базе сведений о языке, учета возможностей машин и здравых представлений о структуре алгоритма (типичная задача кибернетики);
- в) разработка рациональных методов усовершенствования алгоритмов на основе машинных экспериментов (опять задача кибернетики);
- г) выяснение новых требований к изучению языка (здесь переплетаются соображения кибернетики и лингвистики, причем ведущими являются первыми);
- д) разработка методов программирования переводческих алгоритмов и реальное построение программ (также задача кибернетики);
- е) проведение машинных экспериментов и анализ результатов (в основном задача кибернетики, целесообразно участие лингвистов);
- ж) составление словарей с необходимой информацией (идейная часть — структура информации — смешанная задача лингвистики и кибернетики, техническая часть — составление словарей — задача лингвистики);
- з) очерчивание характера текстов, на которые рассчитан перевод (задача кибернетиков).

2. Для выполнения такой сложной задачи, очевидно, требуется большой коллектив. Необходимо обеспечить разумное распределение заданий внутри коллектива и расстановку сил. Для этого необходимо видеть всю задачу в целом. Такие обязанности падают всегда на долю кибернетиков.

3. Уже из предыдущего видно, что руководящую роль, определяющую, в работах по МП должны выполнить математики-кибернетики.

4. Фактически в работе по МП было обнаружено, что существующие в лингвистике способы описания языка не могут служить основой разработки алгоритмов перевода ввиду их недостаточной отчетливости, а также ввиду того, что их использование предполагает априорное понимание смысла текста. Ввиду этого пункт 1а) пришлось очень скоро заменить пунктом 1г), и требования к описанию языка пришлось разрабатывать одновременно с разработкой алгоритмов. После первых серий экспериментов стали проявляться новые требования к описанию языка, а также новые интуитивные представления о том, какими эти описания должны быть. Это вызвало к жизни ряд весьма интересных работ лингвистов (Мельчук, Жолковский, Мартемьянов). Эти новые способы описания языка лучше соответствуют требованиям МП, но, по-видимому, это направление далеко не исчерпано. Складывается представление, что описание языка распадается на ярусы и для каждого типа алгоритмов целесообразно ограничиваться некоторыми определенными ярусами описания языка.

5. В то же время эти описания языка отнюдь не заменяют алгоритмов перевода, тогда как у лингвистов нередко складывается превратное впечатление, что разрабатываемые ими описания языка представляют собой готовые алгоритмы или способы их замены.

6. К сожалению, в настоящее время в области МП лингвисты более активны, чем кибернетики, и нередко оказывается, что математики идут на поводу у лингвистов, вместо того, чтобы определять стратегическое направление этой области. Именно это обстоятельство, наряду с весьма неблагоприятной административной бюрократической конъюнктурой, сильно задерживает реализацию МП (сравнение с ходом дел в Гренобльской группе).

7. Общую стратегию в МП должны определять математики, исходя из стремления к реализации МП. При этом реализация должна вестись поэтапно. Начинать нужно с ограниченных текстов — ограничение области, которое ведет, с одной стороны, к большому облегчению решения вопросов об омоимии, сужению семантики, а в некоторой степени и ограничению синтаксических структур. Кроме того, это позволит начать работу с меньшим объемом словаря, а затем нужно постепенно расширить возможности и, в частности, дифференцировать алгоритмы по областям. Не следует бояться того, что некоторая часть фраз будет переводиться неудачно.

8. Весьма существенно то, что если МП заработает хотя бы в ограниченных пределах, то можно рассчитывать на то, что это повлияет на стилистику научных статей. Как только МП окажется в состоянии ускорять распространение информации,

*) К сожалению, как это неоднократно отмечалось, в том числе и в данной статье, реальная ситуация в лингвистике совершенно не соответствует этому требованию. Слишком многие проблемы и явления из области синтаксиса и семантики языков лингвистами раньше вообще не изучались. Поэтому в действительности работы в области создания алгоритмов МП и продвижение в области изучения естественных языков переплетены очень сложным образом.

авторы и редакторы будут стремиться к тому, чтобы приспособлять язык к возможностям МП.

9. Весьма существенна разработка специальных методов программирования алгоритмов МП с тем расчетом, чтобы внесение исправлений на основании экспериментов было возможно просто доводить до программы.

10. Возникает много вопросов, связанных с приданием алгоритмам МП рациональной структуры — в каких случаях лучше таблицы, а в каких случаях лучше иметь дополнительный кусок программы, какую форму следует придавать отдельным правилам, операторам, группам операторов, а также таблицам. Здесь же возникнут и более глубокие вопросы, контактирующие с областью дискретного анализа, например, с работами Лупанова О. Б. по локальному кодированию и Журавлева Ю. И. по локальным алгоритмам. Нередко полезны соображения шенноновского типа: часто встречающиеся вещи должны делаться особенно экономно и рационально. Весьма возможно, что алгоритмам нужно придавать некоторую иерархическую структуру — сравнительно простой алгоритм действует для всех фраз, некоторый комплекс признаков служит для проверки того, какие из этих фраз переведены приемлемо. В случае отрицательного результата, в зависимости от результатов проверки включается тот или другой алгоритм второго яруса, более сложной структуры и поэтому функционирующий значительно медленнее. Кроме того, эти алгоритмы могут иметь универсальные части и части, приспособленные к специальным областям. Возникающие здесь вопросы могут разрабатываться только математиками.

11. В целом ряде случаев целесообразно конструировать математические модели явлений, с которыми приходится сталкиваться при разработке алгоритмов МП и выбирать ту или другую структуру алгоритмов в зависимости от экспериментов с этими моделями.

В записке, которая приведена выше (см. стр. 63), есть такая фраза: «Нецелесообразность буквального копирования действий человека». Дело в том, что Алексей Андреевич считал методологически ошибочной довольно распространенную точку зрения, трактующую машинный перевод как моделирование деятельности человека, переводящего тексты. Возможности ЭВМ в процессах переработки информации в настоящее время существенно отличаются от возможностей человеческого разума, поэтому перевод при помощи машины означает получение результата, сходного с тем, который получил бы человек (в идеале точно того же), но совсем другими средствами. Алексей Андреевич как-то сказал, что говорить о том, что МП есть моделирование переводческой деятельности человека — так же бессмысленно, как говорить, что автомобиль моделирует лошадь. Подробнее этот вопрос разобран в статье [12]. Поскольку Алексей Андреевич придавал этому вопросу большое значение, мы приведем еще одно его высказывание по этому поводу. В отзыве о докладе «Современное состояние проблемы автоматического перевода» на совещании «Пути автоматизации научно-исследовательских работ» А. А. Ляпунов писал:

Моделирование означает всегда построение некоторой системы, изоморфной некоторой другой данной системе. Когда речь идет о моделировании управляющих систем, перерабатывающих информацию, то предполагается структурно-функциональный изоморфизм моделирующей и моделируемой систем, доведенный до определенной степени глубины членения систем на элементы. Рассматривать машинный перевод как моделирование действий человека-переводчика нельзя, так как ни о каком структурно-функциональном изоморфизме здесь вопрос не ставится.

Зимой 1970/1971 г. Алексей Андреевич продиктовал во время беседы некоторые свои соображения о проблеме МП, которые мы приведем ниже.

Начнем с уточнения постановки задачи МП. В самом общем виде задача может восприниматься так: построить алгоритм и соответствующие ему программы, который может перевести любой текст с одного языка на другой с сохранением всех его особенностей. Такая задача, по крайней мере в настоящее время, должна рассматриваться как недоступная. Основания для такой точки зрения состоят в следующем: каждый алгоритм разрабатывается человеком в некоторый ограниченный отрезок времени. При этом человек использует имеющиеся у него знания, в частности, знание обоих языков. С другой стороны, каждый человеческий язык обладает способностью описать любую совокупность человеческих знаний. Если совокупность человеческих знаний можно представить себе как некоторую формально-логическую систему, то задача пол-

ного анализа любого текста, написанного на некотором языке, средствами того же языка, оказывается в значительной мере аналогична известной задаче Гёделя об установлении непротиворечивости любой достаточно мощной формальной системы средствами этой самой системы. Как известно, в таком виде эта последняя задача неразрешима. С другой стороны, всегда можно построить более широкую формальную систему, в которой непротиворечивость предыдущей системы поддается установлению. В то же время, можно представить себе последовательность все расширяющихся формальных систем таких, что в каждой следующей непротиворечивость предыдущих доказуема.

Рассуждая по аналогии о возможностях машинного перевода или хотя бы о возможности полного анализа произвольного текста, по-видимому, можно себе представить следующую картину. Пусть первый язык, а следовательно, и возможности мышления, описываемые языком, представлены в виде последовательности расширяющихся систем, которые его полностью исчерпывают, но только своей совокупностью. Каждая из этих систем, т. е. ограниченных языков, обладает своей лексикой, морфологией, синтаксисом и семантикой. Будем считать, что каждый ограниченный по объему текст первого языка полностью укладывается в рамки одного какого-нибудь из этих ограниченных языков. Задача каждого алгоритма МП состоит в том, чтобы осуществить перевод любого текста, укладываемого в рамки одного из этих ограниченных подязыков первого языка, на выходной язык. При этом процедура составления алгоритма заведомо требует знания некоторых из более широких подязыков последовательности. Само собой разумеется, что выбор исчерпывающей последовательности подязыков неоднозначен. Трудно указать рациональные принципы построения этих последовательностей, однако реальные работы по МП, по сути дела, строятся, исходя из некоторого интуитивного представления о таких системах. В принципе возможны два пути для построения начальных подязыков. Один из них, в принципе — теоретико-множественный, состоит в очерчивании тех текстов, которые должны быть охвачены алгоритмом. (Алгоритм обязан переводить тексты из указанного множества, он может быть приложен и к текстам вне этого множества без гарантии удачного перевода.) Можно отметить, что уже здесь мы сталкиваемся с вопросом о критериях качества перевода. Эта трудность возникает также и всюду дальше. Будем исходить из следующей формулировки: требуется, чтобы подавляющее большинство людей, владеющих обоими языками и понимающих переводимый текст, признало перевод приемлемым. Второй путь ограничения языка — перечень типов фраз, точнее, структур фраз, которые должны переводиться алгоритмом правильно.

Необходимо отметить чрезвычайную существенность экспериментальной проверки алгоритмов перевода. Сложность и многообразие процесса переработки информации в этих алгоритмах такова, что предусмотреть результат на материале большого объема практически невозможно; также невозможно предвидеть эти результаты, отталкиваясь от теоретических рассуждений грамматической природы. В разработке алгоритмов перевода на начальных этапах был совершенно неизбежен чисто эмпирический процесс построения правил, отталкиваясь на первых порах от традиционной грамматики. В дальнейшем сравнительно скоро выяснилась недостаточность традиционной грамматики для этих целей. Это привело к необходимости построения некоторой новой грамматики, рассчитанной только на выбранный ограниченный язык, но зато поддающийся вполне точному описанию.

В ноябре 1972 г. Алексей Андреевич во время беседы продиктовал две записки — одну о машинном переводе, другую — о проблематике теоретической кибернетики. Ниже приводится первая из них целиком и та часть второй, которая касается непосредственно МП. Здесь более подробно рассматриваются задачи, сформулированные в п. 2 на стр. 64.

Речь идет о рассмотрении алгоритмов, служащих для переработки больших массивов сложно организованной информации, и о способах кодирования сложной и разветвленной информации.

Роль иерархичности.

Во всех случаях имеется такая структура.

1. Имеют место некоторые элементарные акты низшего уровня, которые непосредственно работают над исходной информацией (отдельные чисто локальные правила).
2. Следующий уровень правил использует результаты первичной обработки и т. д.

Правила иерархически классифицируются.

При этом на нижнем уровне число различных структурных типов сравнительно невелико, затем оно растет, потом снова сокращается, на самом верхнем уровне их снова мало, иначе слишком большое число случаев реально не использовалось бы.

(Если говорить об эволюции — более высокие уровни обычно более стабильны.)

Это имеет место в синтезе контактных схем (локальное кодирование — три уровня), в структуре языка как такового и в структуре переводческих алгоритмов.

При изучении алгоритмов и способов кодирования информации целесообразно двигаться с двух сторон — сверху и снизу, где меньше переборов, часто не надо проходить все уровни (морфология — потом фраза, а комбинации по 4—5 слов не просматриваются).

Если присмотреться к локальному кодированию О. Б. Лупанова, то там некоторая группа элементов универсальна и она зависит от природы самих элементов (а не от того, что из этих элементов синтезируется), группа элементов приурочена к реализуемому классу функций, группа элементов приурочена к реализуемой индивидуальной функции.

Очень целесообразно внимательно сопоставить алгоритмы, возникающие в дискретной математике для переработки строго формализованной информации, и алгоритмы МП и вообще большие алгоритмы для переработки неформализованной информации.

Сам факт формализованности информации от особенностей алгоритма, который с ней работает, далек. Алгоритм использует структурные связи, как локальные, так и глобальные, заданные в кодировке информации, причем тоже несколько обобщенно и огрубленно. Вероятно, можно выделить несколько типов структур кодирования информации или информационных связей. Применительно к этим структурам надо устраивать алгоритмы, работающие с этой информацией. При этом характер иерархичности информационных связей имеет много большее значение, чем индивидуальные особенности кодировки. Можно подозревать, что это некоторым образом навязывает структуру анализирующих и синтезирующих алгоритмов, связанных с этой информацией.

Например, синтез контактных схем связан с тем, что схема характеризуется еще и топологией. Вопросы, связанные с кодированием д.н.ф., связаны скорее с тем, что структура д. н. ф. определяется метрическими и комбинаторными соотношениями между гранями куба, где д. н. ф. = 1. Анализирующие и синтезирующие алгоритмы в МП сильно зависят от того, насколько язык синтетический или аналитический. Алгоритмы таких задач как коммивояжера, назначение, расписание — в значительной мере определяются соотношениями между локальными и глобальными связями.

Было бы очень интересно типизировать алгоритмы в зависимости от классов решаемых задач, в частности, порядок просмотра информации, характер неустранимых переборов, каким количеством исходных данных определяется промежуточный результат.

В каждом алгоритме окончательная информация получается из исходной с несколькими промежуточными этапами. Вопрос, сколько ступеней, на какое подмножество исходных данных опирается промежуточный результат. Конечно-разностные схемы, комбинаторные алгоритмы, лингвистические алгоритмы. Поиск эмпирических общих явлений, потом общие соображения, а может быть, аналогии, которые полезно использовать.

Алгоритм МП — это частный, но очень важный случай алгоритмов кодирования, которые встречаются в очень разнообразных задачах кибернетики. Избыточность кодирования языка дает возможность избегать переборов и получать относительно однозначные результаты, по возможности хорошие, причем анализировать приходится тот материал, который дан, поэтому здесь необходим учет разнообразных возможностей. При синтезе целесообразно обыгрывать избыточность для получения таких фраз, которые конструируются возможно единообразно и просто, — при условии правильности. Поэтому, с моей точки зрения, многозначность анализа и многозначность синтеза в МП играют разную роль. Было бы целесообразно иметь некоторые формализованные языки для записи алгоритмов так, чтобы на этих языках можно было бы описывать как алгоритмы над модельными объектами, так и над реальными для того, чтобы родство строения алгоритмов сделать более явным.

В заключение нам хочется привести набросок плана развития работ по МП в нашей стране, сделанный Алексеем Андреевичем осенью 1972 г. Он считал, что имеющийся задел дает достаточную базу для перехода к практической реализации. Алексей Андреевич пытался предпринимать организационные шаги в направлении создания производственного центра по МП, но довести дело до конца не успел.

Приведем сделанный А. А. Ляпуновым набросок плана.

1. К 1975 году будет работать первый вариант МП ФР *).
2. В период 1975—80 гг. МП ФР развивается в производственный.
 - Создать центр по МП (научно-производственный).
 - Разработка первого варианта МП АР (англо-русский).
 - Подготовка вывода, согласованного с набором.
 - Создание специализированных машин.

*) Имелась в виду работа в экспериментальном режиме системы ФР—П, созданной в ИПМ.

3. 1980—90 гг.

ФР совершенствуется в процессе работы.

АР доводится до производственного типа.

Постепенно разрабатывается МП с ряда других языков.

В научных издательствах создается производственный МП и другие организации научно-информационной службы.

Работа на специализированных машинах.

Кадры.

Техника.

ЛИТЕРАТУРА

1. Л я п у н о в А. А., К у л а г и н а О. С., Использование вычислительных машин для перевода с одного языка на другой, Природа 8, 1955.
2. П а н о в Д. Ю., Л я п у н о в А. А., М у х и н И. С., Автоматизация перевода с одного языка на другой, Сессия по научным проблемам автоматизации производства, М., изд. АН СССР, 1956.
3. К е л д ы ш М. В., Л я п у н о в А. А., Ш у р а-Б у р а М. Р., Математические вопросы теории счетных машин, Доклад на сессии АН СССР, 1956.
4. К у з н е ц о в П. С., Л я п у н о в А. А., Р е ф о р м а т с к и й А. А., Основные проблемы машинного перевода, Вопросы языкознания 5, 1956.
5. Л я п у н о в А. А., К у л а г и н а О. С., О работах по машинному переводу Математического института АН СССР, Тезисы Конференции по машинному переводу, М., 1958.
6. Л я п у н о в А. А., Математические исследования, связанные с эксплуатацией вычислительных машин, Сб. «Математика в СССР за 40 лет», 1917—1957, I, Обзорные статьи, М., Физматгиз, 1959.
7. К у л а г и н а О. С., Л я п у н о в А. А., М е л ь ч у к И. А., М о л о ш н а я Т. Н., Теоретические основы машинного перевода на русский язык, Исследования по славянскому языкознанию, М., 1961.
8. Б а г р и н о в с к а я Г. П., К у л а г и н а О. С., Л я п у н о в А. А., М о л о ш н а я Т. Н., Некоторые вопросы математической лингвистики, возникающие в связи с машинным переводом, Сб. «Машинный перевод и прикладная лингвистика», вып. 6, М., Труды МГПИИЯ им. Мориса Тореза, 1961.
9. Л я п у н о в А. А., Я б л о н с к и й С. В., Теоретические проблемы кибернетики, Сб. «Проблемы кибернетики», вып. 9, М., Физматгиз, 1963.
10. Л я п у н о в А. А., Предисловие к книге И. А. Мельчука «Автоматический синтаксический анализ», серия «Кибернетика в монографиях», Новосибирск, изд. СО АН СССР, 1964.
11. Л я п у н о в А. А., Б а г р и н о в с к а я Г. П., К вопросу о программировании машинного перевода. Конференция по машинному переводу. Тезисы докладов, Ереван, 1967.
12. Б а г р и н о в с к а я Г. П., К у л а г и н а О. С., Л я п у н о в А. А., О некоторых методологических вопросах, относящихся к машинному переводу, Сб. «О некоторых вопросах теоретической кибернетики и алгоритмах программирования», Новосибирск, СО АН СССР, 1971.
13. Б а г р и н о в с к а я Г. П., Л я п у н о в А. А., О задачах математической лингвистики, Сб. «О некоторых вопросах теоретической кибернетики и алгоритмах программирования», Новосибирск, СО АН СССР, 1971.
14. Л я п у н о в А. А., Послесловие редактора к статье О. С. Кулагиной «О машинном переводе текстов на естественных языках», Сб. «Проблемы кибернетики», вып. 27, М., «Наука», 1973.
15. М о л о ш н а я Т. Н., Алгоритм перевода с английского языка на русский, Сб. «Проблемы кибернетики», вып. 3, М., Физматгиз, 1960.

Поступила в редакцию 30 VIII 1974