

W.J. Hutchins (Ed.). Early years in machine translation: memoirs and biographies of pioneers (Amsterdam studies in the theory and history of linguistic science. Series III. Studies in the history of language sciences). John Benjamins Publishing Company. Amsterdam; Philadelphia. 2000. xii + 405 p.

В середине прошлого века ведущие авторитеты тогда еще новой структурной лингвистики все настойчивее подчеркивали, что задачей языкознания является (подобно тому, как это имеет место в математике) построение непротиворечивых "красивых" теорий – теорий, которые не нуждаются в экспериментальной проверке своей онтологии. "(...) экспериментальные данные, – писал Л. Ельмслев, – никогда не могут усилить или ослабить теорию, они могут усилить или ослабить только ее пригодность" [Ельмслев 1999: 142]. Аналогичные утверждения встречаются в ранних работах Н. Хомского [Хомский 1965: 250 и сл.]. Схожи с ними рассуждения некоторых ученых и в 60-е гг. [Гладкий, Мельчук

1969: 6]. Однако ход развития науки, как известно, полон неожиданностей и парадоксов. Именно в этот период произошло зарождение и бурное формирование совершенно новой ветви языкознания – инженерной лингвистики, точнее, ее центрального направления – машинного перевода (МП), которому было суждено стать наиболее сильным средством оперативной экспериментальной проверки состоятельности лингвистических теорий.

Зарождению МП и его первым шагам, а отчасти оценке идей и замыслов его пионеров с точки зрения сегодняшней практики автоматической переработки текста (АПТ) посвящен рецензируемый сборник. Организатором и редактором сборника, а также

автором нескольких входящих в него статей является известный английский историк МП Дж. Хатчинс [Hutchins 1986].

Из воспоминаний пионеров читатель узнает, с каким трудом формировалась теория и технология МП. И это понятно. Перед зачинателями МП стояла трудная задача совместить недостаточно упорядоченные представления о строении языка и речи, а также нестрогие приемы анализа и описания текста, с одной стороны, с достаточно четко и последовательно отработанной математической технологией, на которую опиралась быстро развивавшаяся техника алгоритмизации и программирования, с другой. Каких-либо серьезных прототипов для решения этой задачи в конце 50-х – начале 60-х гг. не существовало.

Сложность ситуации усугублялась тем, что романтическая идея создать говорящую и переводящую машины привлекла специалистов самых разнообразных специальностей. Здесь наряду с лингвистами [Э. Рейфлер, США (Micklesen: 21)¹; А. Браун, США (Brown: 134); В. Леманн, США (Lehmann: 164); С. Лэм, США (Lamb: 195) и др.], библиотекарями [Дж. Хатчинс, Великобритания (Hutchins: 15)], математиками [У. Уивер, США (Hutchins: 17); П. Сгалл, Чехия (Kirschner: 349–350); А. Людсканов, Болгария (Paskaleva: 364–365) и др.], инженерами, электро- и радиотехниками [Г. Кинг, США (Hutchins: 171); М. Гросс, Франция (Gross: 330); Э. Бут, Великобритания (Booth: 260–261), Х. Вада, Япония (Wada: 385) и др.] не только вращались, но пытались руководить исследованиями журналисты [Э. фон Глазерсфельд, США (Glaserfeld: 324)], деятели Коминтерна и переводчики СМЕРШ'а [В.Ю. Розенцвейг – брат известной советской разведчицы Е.В. Зарубиной-Розенцвейг (Mel'čuk: 208)].

Лишь немногие из них, как например Й. Бар-Хиллель [Израиль] (Hutchins: 299), П. Тома [США] (Toma: 145), А. Эттингер [США] (Oettinger: 73), В. Ингве [США] (Yngve: 72), обладали одновременно наряду с физико-математическим или инженерным образованием знанием разных языков, или сочетали профессиональную лингвистическую подготовку с некоторыми математическими или инженерными знаниями. Среди последних М. Васконселлос [США] (Vasconcellos: 96), М. Заречник [США] (Zarechnak: 128), а также отечественные ученые Н.Д. Андреев, Р.Г. Пиотровский (Piotrowski: 234–242), Г.Г. Белоголов, Ю.Н. Марчук (Marčuk: 251)².

¹ В круглых скобках даются ссылки на авторов статей и страницы рецензируемого сборника.

² К сожалению, не упоминаются имена

Поэтому одной из основных трудностей в развитии МП было преодоление различий в подходе к структуре языка филологов, с одной стороны, и математиков и техников, с другой. Эти различия в подходах быстрее сглаживались в творческих "тандемах" филолог – математик или филолог – техник, примером которых могли служить такие содружества московских ученых, как И.К. Бельской с Д.Ю. Пановым (Mološnaia: 228; Marčuk: 244–245), И.А. Мельчука с О.С. Кулагиной и А.В. Гладким (Mel'čuk: 212–216).

Трудности в достижении научного взаимопонимания стали причиной появления в 50–60-е гг. ряда альтернативных подходов, из которых немногие выдержали проверку временем.

В 50-е гг. умы пионеров МП занимали идеи полностью автоматического высококачественного перевода (fully automatic high quality translation – FАНQT), выросшие из постулатов всеобщей математики Р. Декарта – Г.В. Лейбница и логики Б. Рассела и опиравшиеся на предположении Л. Ельмслева и Н. Хомского о том, что естественный язык, так же, как искусственные языки математики или логики, представляет собой исчисление (Hutchins: 3, 8–9; Oettinger: 80, Montgomery: 97–109; Kulagina: 197–203; Piotrowski: 234–235). Идея высококачественного перевода, ассоциируемого с разработкой логического универсального метаязыка, стала центром теоретических изысканий в области АПТ. Порой автоматический перевод представлялся в несколько парадоксальной форме – "перевод без перевода, без машин, без алгоритмов" [Мельчук, Равич 1967: 8–9].

К концу 50-х годов стал очевидным нарастающий разрыв между теоретической и практической ориентациями на FАНQT и реальные рабочие системы, выполняющие узкие задачи. "Алгебраическая лингвистика", опиравшаяся на идею полной формализации языка, оказалась слишком непродуктивной. Главной неразрешимой проблемой оставалось распознавание семантики [Дрейфус 1978]. Требовалось решить вопрос расширения и формализации словарей, изучить различия в синтаксической структуре языков и разработать принципы синтаксического анализа. Стало вполне очевидно, что создание метаязыка или логического описания языка, сопровождаемого формализацией человеческих знаний и психических процессов, потребует нескольких десятков лет работы сотен лингвистов, психологов, математиков и програм-

таких ученых, как А. Есенин-Вольпин, Д. Лахути, В. Чернявский [Pevzner 2001: 81–86], внесших заметный вклад в становление МП.

мистов (Piotrowski: 235; Marčuk: 245–248). Закономерным следствием творческого тупика "алгебраической лингвистики" стало собрание в октябре 1963 г. Консультативного комитета по автоматической переработке естественно-го языка (automatic language processing advisory committee – ALPAC) с целью изучить состояние дел в области МП, что привело в 60-е гг. к откату в финансовой поддержке МП в США.

Значимыми событиями 50-х гг. стали демонстрации Джорджтаунской системы русско-английского МП для нужд Пентагона, – системы, ориентированной не на FANQT, а на грубую с точки зрения синтаксиса, семантики и стилистики, но понятную для потребителя и устойчивую переработку текста (Vasconcellos: 90–94; Zarechnak: 111–127). Эти опыты стимулировали выделение значительных средств от Национального научного фонда, а затем и от таких Федеральных служб США, как ЦРУ, военно-воздушные, сухопутные и военно-морские силы и др. Благодаря этой поддержке активизировали свою деятельность и другие американские коллективы. Среди них группы под руководством В. Ингве (Массачусетский технологический институт); А. Эттингера (Гарвардский университет); П. Тома (Калифорнийский технологический институт), присоединившегося позднее к Джорджтаунской группе, возглавляемой Л. Достертом; С. Лэма (Калифорнийский университет в г. Беркли) и др.

Проверкой реально работающих систем является реализация различных форм АПТ на текстах, взятых наугад. Уже на рубеже 50–60-х гг. американские исследовательские группы не только организовали показы "грубого" МП, но осуществили массовый пословно-пооборотный перевод русских документов по технической, медицинской, политической и др. тематике [LES 1960: 185–348].

Деятельность пионеров послужила импульсом для исследований в Чехословакии, Болгарии, Италии, Франции, Японии, Китае, Англии и др. странах (Booth: 253–261; Sparck Jones: 263–278; Wilks: 279–297; Glasersfeld: 313–324; Gross: 325–330; Kirschner: 349–360; Paskaleva: 361–376; Wada: 377–385). Особо заслуживают внимания работы советских коллективов, которые были среди первых разработчиков МП. Так, в 1954–1955 гг. сразу после знаменитого эксперимента в Джорджтаунском университете по созданию русско-английской системы перевода к МП обратились ученые из Института точной механики и вычислительной техники (ИТМиВТ) Д.Ю. Панов и И.К. Бельская. Они первыми дали относительно понятный перевод текста, заранее введенного в машину вместе с программой его преобразования.

К сожалению, малочисленность и слабость отечественной вычислительной техники тормозили создание работающих систем МП. Первый в СССР опыт по машинной переработке английских и немецких наугад взятых текстов был осуществлен только в конце 60-х гг. в Ленинграде группой "Статистика речи" (СтР), руководимой профессором Р.Г. Пиотровским (газета "Ленинградская правда" 30.05.68).

В 60-е гг. зарождается новое направление научно-исследовательской деятельности зарубежных и русских ученых. Преодолев наивно-романтическую мечту о "высококачественном автоматическом переводе", которая подогрела работы по созданию систем АПТ первого поколения, отечественная инженерная лингвистика и некоторые американские коллективы направили свои усилия на разработку систем АПТ и МП, способных выдавать грамматически и стилистически не вполне корректный, но отражающий основное содержание обрабатываемого документа перевод. Среди них Джорджтаунская система, продемонстрированная в декабре 1961 г. в ЕВРАТОМ'е (Европейское сообщество по атомной энергии) (Испра, Италия); система SYSTRAN (акроним для "system of translation") П. Тома, разработка которой началась в 1964 г. в Германии (в 1974 г. система SYSTRAN активно использовалась Национальным управлением по авионавигации и исследованию космического пространства – NASA – в период реализации проекта "Аполлон-Союз"); разрабатываемая с 1963 г. в Советском Союзе англо-русская система для перевода патентной документации, использовавшаяся в Центральном бюро патентования в Москве; система CULT для перевода математических текстов с китайского языка, созданная в 1972 г. в Китайском университете в Гонконге; первая полностью автоматическая система для перевода сводок погоды, разработанная в Центре автоматической переработки лингвистических данных (Centre de traitement automatisé des données linguistiques – CETADOL) и введенная в практику в 1976 г. в Монреале и др. (Hutchins: 12, 303; Montgomery: 98; Brown: 132; Lehmann: 158).

В 70-е гг. активизировалась деятельность группы СтР. Она включала лингвистов и преподавателей языка, математиков и программистов, психологов и психиатров. Головной коллектив группы работал в Ленинграде, филиалы находились в Минске, Кишиневе, Риге, Тарту, Харькове, Баку, республиках Средней Азии, а позднее в США, Франции, Голландии, Израиле и др. странах.

Около 40 лет в группе решаются проблемы АПТ, в том числе МП, а также линг-

вистические и лингводидактические задачи искусственного интеллекта. Группа СтР подошла к задаче компьютеризации лингвистики самым, казалось бы, примитивным образом, то есть через выделение наиболее информативных и частотных слов, выражений и грамматических структур. Они и должны были составить ядро языковых знаний и умений компьютера, позволяющих ему решать не слишком сложные лингвистические задачи. Этот упрощенный прагматический подход, объединяющий теорию и практику использования лингвистических автоматов (ЛА)³, получил название "инженерная лингвистика" (на Западе в конце 80-х гг. этот термин вошел в употребление в виде английского *language engineering*).

Важным моментом в научно-исследовательской деятельности группы является то, что она начала гораздо раньше, чем американцы, работать над вопросом извлечения с помощью компьютера основного смысла документа. Задачу извлечения основной идеи документа решает так называемое машинное реферирование. Параллельно с МП группа в конце 60-х гг. занялась построением программ такого реферирования. Результатом работы стало создание в середине 70-х гг. программы китайско-русского МП, ориентированной прежде всего на перевод почтовых телеграфных сообщений. За эту компьютерную систему в 1982 г. группе СтР была присуждена премия Совета Министров.

В дальнейшем выделившиеся из группы СтР научно-исследовательские коллективы в России и Украине, Молдове и Белоруссии выпустили на Европейский рынок ряд программ АПТ (MULTIS-SILOD, STYLUS, SARMA, PARS и др.), осуществляющих реальную коммерческую обработку деловых текстов (Piotrowski: 237; Marčuk: 247–249).

Материалы статей сборника позволяют сделать вывод о том, что неудачи попыток создания высококачественного МП и использования языка-посредника были связаны прежде всего с нечеткой ассоциативной природой естественного языка (Piotrowski: 237–239). Одним из первых ученых, еще в начале 50-х гг. обративших внимание на различие между искусственным и естественным языками и изложивших свои сомнения по поводу создания FANQT, был известный израильский философ-логик Й. Бар-Хиллель (Hutchins: 9, 309–311; Montgomery: 97).

³ Лингвистический автомат представляет собой сочетание компьютера, лингвистического алгоритма и реализующей его программы (программное и информационно-лингвистическое обеспечение).

Пренебрежение различиями и парадоксами, отделяющими естественный язык от языка компьютера [Заде 1976; Мельников 1978: 218 и сл.], неизменно заводило и заводит в тупик всех разработчиков систем АПТ и МП, которые пытались и пытаются работать в русле дедуктивно-логической стратегии. За более чем 40 лет существования проблемы АПТ ни одному из европейских, советских (российских), американских коллективов, работавших в этом ключе, не удалось построить реально работающую систему МП. Поэтому уже в середине 60-х гг. ряд западных и российских коллективов – в первую очередь международная группа СтР – обратились к поиску альтернативной стратегии и выработке новой инженерно-лингвистической технологии.

Такой стратегией является информационно-статистический и многоуровневый подход к решению задачи АПТ. Одной из ключевых идей развития указанной методологии является имитация речевого поведения человека. Последняя состоит в пошаговом сокращении неопределенности, в которой оказывается ЛА, – неопределенности, порождаемой, с одной стороны, многозначностью словарных лексических единиц, морфологических форм и синтаксических схем, содержащихся в тексте, а с другой – недостатком лингвистических и энциклопедических знаний.

Сокращение неопределенности идет снизу вверх от лексических фактов к семантико-синтаксической и прагматической цели. Неопределенность снимается благодаря узкой направленности машинно-переводческого модуля на конкретную предметную область (ПО) и тщательному отбору терминов и терминологических словосочетаний в автоматические словари (Piotrowski: 240–241).

Выясняется, что одной из сложнейших задач системы МП является устранение многозначности. Оказывается, целый пласт терминологических слов и словосочетаний не может обойтись без вторичного означивания (вторичного семиозиса). Причем эта "терминологическая болезнь" характерна как для терминов, относящихся к различным ПО, так и обслуживающих одну ПО. Таким образом, язык терминологии несмотря на предъявляемые к ней требования однозначности, стилистической нейтральности и контекстуальной независимости не является формальным исчислением. Выдаваемый ЛА перевод текста обычно далек от стилистического совершенства, а значит стопроцентное преодоление барьера отторжения, отделяющего естественный язык от языка ЭВМ, является делом будущего.

Итак, противоречивая и полная драматических коллизий история МП служит убедительным доказательством того, что состоятельность лингвистических теорий должна подвергаться экспериментальной проверке. В этом и состоит большое научное значение рецензируемого сборника.

СПИСОК ЛИТЕРАТУРЫ

- Гладкий А.В., Мельчук И.А. 1969 – Элементы математической лингвистики. М., 1969.
- Дрейфус Х. 1978 – Чего не могут вычислительные машины. Критика искусственного разума. М., 1978.
- Ельмслев Л. 1999 – Прологомены к теории языка // Зарубежная лингвистика. I / Общ. ред. В.А. Звегинцева и Н.С.Чемоданова. М., 1999.
- Заде Л. 1976 – Понятие лингвистической переменной и его применение к принятию приближенных решений. М., 1976.

- Мельников Г.П. 1978 – Системология и языковые аспекты кибернетики / Под ред. Ю.Г. Косарева. М., 1978.
- Мельчук И.А., Равич Р.Д. 1967 – Автоматический перевод. 1949–1963 Критико-библиографический справочник / Под ред. Г.С. Цвейга и Э.К. Кузнецовой. М., 1967.
- Хомский Н. 1965 – Объяснительные модели в лингвистике // Математическая логика и ее применения / Под ред. Э. Нагеля, П. Санпса и А. Тарского. М., 1965.
- Hutchins W.J. 1986 – Machine translation: past, present, future. Chichester, 1986.
- LES 1960 – Linguistic and engineering studies in the automatic translation of scientific Russian into English // Technical Report. Phase II. Seattle, 1960.
- Pevzner B. 2001 – First steps of language engineering in the USSR: The 50s through 70s // Machine translation / Ed. by Dr M. S. Blekhan. New Delhi, 2001.

Н Ю. Зайцева