***Р.Г. Пиотровский.* Лингвистический автомат и его речемыслительное обоснование**. Минск: МГЛУ, 1999. 196 с. ***Р.Г. Пиотровский.* Лингвистический автомат (в исследовании и непрерывном обучении).** Санкт-Петербург: Изд-во РГПУ, 1999. 256 с.

В 1999 и 2000 гг. вышли из печати две новые книги проф. Р.Г. Пиотровского, основавшего в 1959 г. и руководившего в 1959–1991 гг. общесоюзной группой "Статистика речи" (СтР), а в настоящее время возглавляющего международный коллектив с этим же названием (члены этого коллектива работают в России, Беларуси, Республике Молдова, Израиле, ФРГ, Канаде и других странах). В двух дополняющих друг друга книгах – "Лингвистический автомат и его речемыслительное обоснование" (далее ЛА-1) и "Лингвистический автомат (в исследовании и непрерывном обучении)" (далее ЛА-2) – суммируется более чем сорокалетний опыт работы самого автора, а также его коллег и учеников в области лингвистического моделирования и семиотики (ЛА-1, гл. 1–3; ЛА-2, гл. 1–2), информатики и вычислительной техники (ЛА-1, гл. 5–7; ЛА-2, гл. 4–7), нейро- и психолингвистики (ЛА-1, гл. 4; ЛА-2, гл. 3), искусственного интеллекта (ИИ), статистико-комбинаторного описания речи и автоматической переработки текста (АПТ), в том числе машинного перевода (МП) и компьютерной оптимизации преподавания языков (ЛА-1, гл. 6–8; ЛА-2, гл. 6–9).

Междисциплинарное разнообразие тематики не случайно. Оно объясняется необходимостью детально познакомить читателя с теми концепциями: общелингвистической (ЛА-1, гл. 1–3; ЛА-2, гл. 1–2), психолингвистической (ЛА-1, гл. 4; ЛА-2, гл. 3) и информационно-статистической (ЛА-1, гл. 5–7; ЛА-2, гл. 4–7), – на базе которых строились в группе СтР реально действующие системы МП (МУЛТИС/SILOD) и на которые опираются современные коммерческие отечественные системы МП (ПРОМТ/СТАЙЛОС, САРМА/СОКРАТ, ПАРС).

Отличаясь, главным образом, разной степенью представления прикладного аспекта, книги объединены теоретической базой, которой посвящены главы, рассматривающие те вопросы в науке о языке и смежных дисциплинах, последовательное решение которых могло бы помочь в преодолении эпистемологических и технологических барьеров, встающих на пути моделирования речемыслительной деятельности человека (РМД) и создания ее компьютерных аналогов – систем ИИ.

Центральной эпистемологической и онтологической проблемой, рассматриваемой в работах, является соотношение природы естественного и искусственного языков. Солидаризуясь с такими учеными как Л. Заде, Г. Дрейфус, Г.П. Мельников, В.В. Налимов, автор утверждает, что

- ЕЯ имеет нечетко-множественное построение, применяющее нечеткую логику и отношение толерантности между его элементами (ЛА-1, с. 6–22; ЛА-2, с. 5–9, 27–34),
- ЕЯ не является исчислением, но представляет собой открытую коммуникативную систему, которая не только применяет узуальные, то есть социально закрепленные связи между означающим и означаемым определенного знака и между различными знаками, но также использует "творческий" вторичный семиозис, или окказиональные ассоциации между обоими компонентами знака, и не предусмотренные правилами логического исчисления ненормативные соотнесения

знаков (ЛА-1, с. 6–9, 35–109; ЛА-2, с. 5–9, 43–126),

- эпистемологическое допущение, согласно которому речемыслительная деятельность человека, как и все его разумное поведение, может быть полностью формализована в терминах единой системы эвристических правил, неверно (ЛА-1, с. 6–9, 23–34, 110–172; ЛА-2, с. 5–42, 126–239).

Опыт построения работающих систем МП показал, что между языком компьютера и языком человека существует барьер отторжения между описаниями стационарных процессов в неживой материи (ср. "язык" компьютера), с одной стороны, и описанием нестационарных процессов, характерных для живой природы, в том числе для лингвистического поведения человека, с другой. Этот барьер реализуется в таких генетических парадоксах человека и лингвистического робота, как

- противоречие между открытым, вечно развивающимся ЕЯ и закрытым, не допускающим самопроизвольного изменения и развития "языком" ЭВМ,
- несоответствие между эквивалентной природой множеств компьютерного языка и толерантностью лингвистических множеств,
- противоречие между единственным для компьютера смыслом текста и его единиц и многоаспектностью естественно-речевого сообщения, несущего обычно три типа смыслов, из которых два диктуются прагматикой коммуникантов (речь здесь идет об индивидуальных авторском и перцептивном смыслах), а третий является независимым от них социологизованным, коллективным смыслом.

Автор напоминает, что пренебрежение указанными различиями и парадоксами неизменно заводило в тупик тех разработчиков систем МП, которые упорно продолжали работать в русле "алгебраической" стратегии, рассматривающей ЕЯ подобно искусственным языкам математики, программирования и т.п. как исчисления, надеясь путем все более углубляющейся и охватывающей новые языковые уровни формализации получить на компьютере высококачественный смысловой перевод. В результате их работа над созданием реально функционирующих систем МП, автоматического индексирования, реферирования и других видов смысловой переработки текста парализовывалась и превращалась в погоню за горизонтом.

Само осознание различий, отделяющих ЕЯ от языка лингвистического автомата, еще не снимает, разумеется, всех трудностей, возникающих при построении систем МП (эти системы не могут быть реализованы иначе, чем на компьютерном языке-исчислении). Поэтому в рецензируемых работах рассматриваются нетривиальные теоретические подходы, которые могли бы ослабить отторжение ЕЯ со стороны компьютерного языка-исчисления.

В частности, одной из основных задач, связанных с компьютерным моделированием РМД человека, является, с одной стороны, формальное описание этой нечеткости и континуальности (ЛА-1, гл. 1; ЛА-2, гл. 1), а с другой – поиски путей представления нечетких и размытых семиотических объектов РМД в виде четких и дискретных аналогов (ЛА-1, гл. 3, 4; ЛА-2, гл. 2, 3), которые могут быть восприняты и переработаны ЛА (ЛА-1, гл. 6, 7; ЛА-2, гл. 6, 7, 8). Отметим, что термин "знак" трактуется автором в русле соссюрианской традиции как билатеральная психическая сущность. Это дает бóльшие возможности по сравнению с "треугольником" Пирса-Фреге-Морриса при описании природы указанного выше барьера отторжения.

Задачу снижения барьера, отделяющего ЕЯ от "языка" компьютера, трудно решить без понимания существа синергетических механизмов языка и речи (эта проблематика интенсивно разрабатывается международным коллективом, руководимым немецкими учеными Г. Альтманном и Р. Кёлером). Поэтому Р.Г. Пиотровский уделил особое внимание синергетическим механизмам РМД в индоевропейских и некоторых урало-алтайских языках (тюркских, финно-угорских) и возможностям их моделирования (ЛА-1, гл. 2, 5, 8). Основываясь на экспериментальных данных и, в первую очередь, информационно-статистических измерениях текстов моделируемых разноструктурных языков, автор обращает внимание читателя на те аспекты их саморегуляции, учет которых помогает понять причины неудач, постигших многие коллективы разработчиков систем ИИ и других форм АПТ. По ходу дела автор дает описание синергетических механизмов в диахронии индоевропейских языков, сравниваемой эпизодически с историческим развитием других языковых семей. К сожалению, автор не учел наблюдения над механизмами саморегулирования и самоорганизации в большей части языков, оставшихся за рамками рецензируемого сочинения и заслуживающих не меньшего внимания. В этой связи хотелось бы высказать автору пожелание расширить

языковое поле синергетических исследований[1].

В книге ЛА-2 автор сосредотачивает свое внимание на анализе результатов по компьютерному решению лингвистических (включая информационно-статистический подход, гл. 4–7) и лингводидактических задач (гл. 8). Здесь представлены реально работающие системы МП, аннотирования и реферирования текста, а также попытки компьютерного моделирования основных аспектов РМД человека.

Автор показывает, что "интеллектуальные" системы переработки текста должны в будущем функционировать так, как работают сложные системы в информационно-энергетическом пространстве. Это значит, что ЛА и обучающий ЛА (ОЛА) должны строиться не только в виде полифункциональных программ, способных осуществлять различные изолированные операции (грубый лексико-грамматический МП, автоматическое индексирование и аннотирование текста, автоматическое распознавание сканированных письменных текстов, выполненных различной графикой, в том числе иероглификой, распознавание и "понимание" устной речи – ЛА-2, гл. 6–8). ЛА и ОЛА будущего должны уметь с помощью аналога коммуникативно-прагматического оператора (КПО) сопоставлять результаты этих операций, оценивать их адекватность замыслу отправителя и ожиданию получателя информации, а в идеале и учитывать изменение внешних условий человеко-машинного общения, другими словами – отличаться творческим подходом в решении сложных "интеллектуальных" задач (ЛА-2, гл. 9). Это значит, что машинный КПО должен уметь самостоятельно перестраивать в ЛА и ОЛА структуру целей и динамически менять путь переработки текста в пространстве лингвистических задач. Эта перестройка предусматривает изменение приоритетов целей. Цели, считающиеся недостижимыми, откладываются. Например, если автомату в ходе анализа данного фрагмента недостает семантико-синтаксических ресурсов, он должен ограничиться его пословно-пооборотным переводом. При этом отложенные цели не забываются и при необходимости могут актуализироваться. Так, после семантико-синтаксического перевода оставшейся части предложения или абзаца ЛА может вернуться к анализу первого примитивно переведенного фрагмента с тем, чтобы, используя новую информацию, попытаться перевести его.

Следует заметить, однако, что при описании "интеллектуальных" систем переработки текста автор недостаточно учитывает опыт эксплуатации отечественных коммерческих систем МП ПРОМТ и СОКРАТ, а также лучших зарубежных систем АПТ SYSTRAN и GLOBALINK, которые опираются на лингвистическую методологию, близкую описанным в рецензируемых книгах идеям. Так, поиски этими коммерческими коллективами инженерно-лингвистических и алгоритмических путей улучшения МП показали, что, хотя современные промышленные системы МП и ориентированы своими словарями и грамматиками на определенные предметные области, эти системы не способны "понять" его общий смысл и композицию. Они не распознают в документе иноязычные вставки и фрагменты, относящиеся к другим предметным областям. Не воспринимают они и переходы от одной частной тематики внутри основной предметной области к другой. Эти системы механически и бездумно переводят иноязычный текст, не имея возможности контролировать осмысленность своего перевода. Наблюдения показывают, что неограниченное расширение лингвистической базы данных (ЛБД) и "латание" алгоритма и программы приводит чаще всего к росту информационных помех и ухудшает качество перевода. Заметное улучшение качества МП требует замены традиционных статических ЛБД, применяемых в современных системах, на динамические базы знаний, которые давали бы возможность лингвистическому автомату преодолеть три операционных ступени на пути к адекватному "пониманию" и переводу текста, а именно:

1) ступень его нормализации, то есть распознания его формальной и смысловой организации,

2) ступень адекватного перевода ("понимания") отдельных составляющих текста (слов, словосочетаний и отдельных простых предложений),

3) ступень "понимания" текста в целом, позволющая скорректировать ошибки, допущенные ЛА на первых двух ступенях переработки текста.

Необходимость такого трехступенчатого построения новой системы МП стала особенно очевидной в ходе создания в 1997–2000 гг. группой СтР экспериментальной системы устного машинного перевода ORAL SILOD (см. [Beliaeva, Zaitseva et al. 2000]).

[1] В частности, следовало бы использовать такие источники, как [КТААЯ 1982; Мельников 1968; Мельников, Охотина 1971; Breiter 1994; Nettle 1998].

Описание узловых проблем ИИ, представленное в рецензируемых работах, привлечет внимание ученых разных направлений, а особенно специалистов в области теоретического и прикладного языкознания, информатики и вычислительной техники. Оба сочинения могут быть использованы в качестве учебных пособий для студентов, аспирантов и преподавателей университетов самых разных профилей, которые предполагают заниматься решением разных лингвистических задач, включая в первую очередь задачи АПТ и компьютерной оптимизации преподавания родного и иностранного языков, предусматривающие творческое взаимодействие учащегося с многоцелевой автоматизированной лингводидактической системой. Материал обеих книг может быть использован в преподавании курсов "Языкознание", "Введение в филологию", "Контрастивная лингвистика", "Естественный язык и язык компьютера", "Языкознание и математика", "Математика и информатика", "Основы теории перевода", "Лингвистические автоматы", "Компьютерные обучающие программы", "Технические и аудиовизуальные средства обучения (информационно-педагогические технологии)".

Оценивая сбалансированность архитектоники обеих книг, хотелось бы заметить следующее. В книге "Лингвистический автомат (в исследовании и непрерывном обучении)", учитывая ее бóльшую прикладную ориентацию, можно было бы расширить описание статистической методики, подробнее описать использование ЛА в обучении родному и иностранному языкам, ввести контрольные вопросы и упражнения, а также выделить особо в библиографии работы учебного характера, доступные широкому кругу учащихся.

В заключение следует подчеркнуть, что лингвистические аспекты искусственного интеллекта и автоматической переработки текста обсуждаются в монографиях лаконично, без фразерства, с лингвистической точностью, фактологической и библиографической эрудицией, принципами объективности и краткости, стилистическим изяществом.

СПИСОК ЛИТЕРАТУРЫ

КТААЯ 1982 – Квантитативная типология афро-азиатских языков / Отв. ред. В.Б. Касевич, С.Е. Яхонтов. Л., 1982.

Мельников Г.П. 1968 – Системный анализ причин своеобразия семитского консонантизма (методические разработки). М., 1968.

Мельников Г.П., Охотина Н.В. 1971 – Выявление детерминанты в классификации морфем банту (на материале суахили) // Проблемы африканского языкознания. Типология, компаративистика, описание языков. М., 1971.

Beliaeva L., Zaitzeva N. et al. 2000 – Oral SILOD – an experimental system of oral machine translation // Speech and computer. SPECOM'2000 / International Workshop: Proceedings (25–28 September 2000). St.-Petersburg. 2000.

Breiter M.A. 1994 – Length of Chinese words in relation to their other systemic features // Journal of quantitative linguistics. V. 1. 1994. № 3.

Nettle D 1998 – Coevolution of phonology and the lexicon in twelve languages of West Africa // Journal of quantitative linguistics. V. 5. 1998. № 3.

Н.Ю. Зайцева, Ю.А. Косарев