

Прометей науки.

Слово об А.С. Нариньяни

Загорулько Ю.А., кандидат технических наук

Кононенко И.С., научный сотрудник



Ушёл из жизни Александр Семёнович Нариньяни — необыкновенно харизматичный, творческий и обаятельный человек. Для своих учеников и соратников он всегда был неиссякаемым источником идей, оптимизма и веры в лучшее завтра.

Область научных интересов Александра Семёновича была весьма обширной — интеллектуальные технологии, компьютерная лингвистика, представление знаний. И в каждой из этих областей он оставил заметный след. Александр Семёнович — основатель нового направления в компьютерных науках: недоопределённых вычислений. Десятки его учеников стали работать в этом направлении, развивать его, создавать практические разработки. По этой тематике было защищено около десятка кандидатских диссертаций.

Александр Семёнович — это человек энциклопедических знаний, патриот, прекрасный собеседник, оратор, способный увлечь своими идеями и повести людей за собой, юморист и душа компании. Именно таким он нам запомнился и таким навсегда останется в нашей памяти.

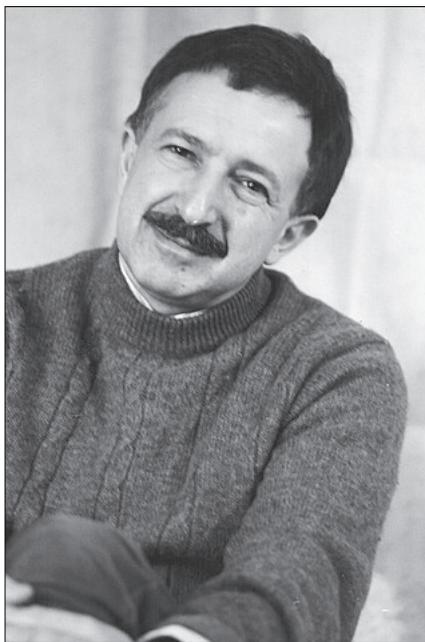
Загоруйко Ю.А., Коноенко И.С.

Прометей науки. Слово о А.С. Нариньяни

Александр Семёнович Нариньяни родился 2 ноября 1937 г. в Москве. В 1956 г. он поступил в Московский инженерно-физический институт. В начале 1962 г. он приехал на преддипломную практику в Новосибирский Академгородок, где и сформировался как учёный.

В начале научной карьеры Александр Семёнович посвятил себя только зарождавшейся тогда теории параллельных вычислений. Здесь им были получены существенные результаты по теории программирования параллельных процессов, доказана принципиальная возможность автоматического распараллеливания последовательных алгоритмов. По этой теме он защитил в 1971 г. кандидатскую диссертацию.

Таким образом, в искусственный интеллект и компьютерную лингвистику Александр Семёнович пришёл из теории параллельных вычислений. По-видимому, отсюда его сверхвысокий параллелизм: он всю жизнь одновременно занимался сразу несколькими проектами (порой число таких проектов доходило до десятка) из абсолютно разных областей знаний — от обработки текстов на естественном языке до моделирования экономики региона и страны.



В начале 1973 г. А.С. Нариньяни возглавил научно-исследовательскую группу по тематике искусственного интеллекта (ИИ), которая была создана в ВЦ СОАН СССР и позже выросла в большую лабораторию. Первоначально эта группа занималась разработкой математического обеспечения для макета шагающего автомата. Чуть позже центральной тематикой группы стала проблема взаимодействия с ЭВМ на естественном языке. В 1974 г. был запущен первый проект по этой теме, получивший название РИТА (Рисунок-Информация-Текст-Автор). Идейным вдохновителем и соруководителем этого проекта был выдающийся лингвист И.А. Мельчук. В рамках этого проекта создавалась экспериментальная система, которая должна была переводить словесное описание в рисунок. Однако на примере такой достаточно простой системы планировалось смоделировать процессы обработки информации на всех уровнях — от линейного текста до его семантического представления и отображения последнего в рисунок.

Далее последовали другие работы в области лингвистики, направленные как на развитие теории (проект «Формальная модель русского языка»), так и на создание прикладных систем (проект ЗАПСИБ).

Целями проекта «Формальная модель русского языка», начатого в 1976 г., были поиск и выбор адекватных средств и структурных решений для формального описания всего спектра языковых явлений. Были заявлены основные идеологические принципы разрабатываемой модели, которые звучали достаточно революционно для того времени: независимость модели от механизмов реализации, сочетание аппаратов зависимостей и составляющих («расширенная составляющая», или компонента, как базовая формальная конструкция), отказ от традиционных уровней представления информации [1]. В рамках этого проекта проведены исследования по формальному представлению различных русских языковых конструкций: наречные и глагольные группы, конструкции с прилагательным и количественные конструкции. Наиболее подробно были описаны числовые параметрические конструкции русского языка, причём исходной точкой лингвистического исследования был принят смысл, которому сопоставлялись различные формы языкового выражения.

Примерно в 1982 г. на семинаре «Диалог» Е.В. Падучева привлекла внимание прикладных исследователей к вопросам языковой прагматики. С новой точки зрения общение, в том числе общение с ЭВМ, является одним из видов человеческой деятельности, имеющей определённые цели и реализующей соответствующие этим целям планы участников общения. А.С. Нариньяни живо откликнулся на эти идеи и в течение довольно значительного периода, пока в лаборатории ИИ продолжались теоретические исследования, формальная модель языка рассматривалась шире — как формальная модель языковой деятельности, в рамках создания которой значительное внимание уделялось моделированию речевого акта [2].

Целью прикладного проекта ЗАПСИБ (ЗАПрос к Справочно-Информационной Базе) было создание технологии построения естественно-языковых процессоров на основе семантически-ориентированного анализа текста [3]. Исходя из идей, опробованных в этом проекте, была построена система InBase, с помощью которой можно создавать для произвольной базы данных интерфейс, «понимающий» произвольные запросы на естественном языке и обеспечивающий прямой доступ к данным для непрофессионального пользователя — например, лица, принимающего решения. В сочетании с системой распознавания речи такой интерфейс может использоваться в системе, в которой требуется «понимание» устных запросов или команд пользователя.

Александр Семёнович считал, что радикальный прогресс в области создания ЕЯ-интерфейсов тесно связан с технологией распознавания голоса. Он отмечал, что интеграция фонетического распознавания голосового сообщения с анализом его содержания должна качественно решить проблему создания массового ЕЯ-интерфейса с машиной: понимание текста на порядок повысит качество средств распознавания речи, а переход от уровня печатного текста к разговорной речи сделает взаимодействие с компьютером на естественном языке по-настоящему естественным.

Если в проекте InBASE принципиально решалась проблема понимания естественно-языковых запросов, коротких сообщений и директив в виде простых приложений, то переход к связному тексту размером от 10 до 20 и более фраз представляет собой сложную задачу даже для ограниченных предметных областей. Экспериментальные исследования по анализу короткого связного текста были начаты в рамках проекта AURA («Автоматическое понимание текстов в ограниченной предметной области»). Технология AURA была ориентирована на широкое применение в делопроизводстве, медицине, образовании, коммерческой деятельности и многих других областях, включая анализ потоков сообщений в деятельности органов государственного управления и правопорядка и в СМИ. На её основе были созданы системы для анализа врачебных описаний результатов рентгенологических исследований грудной клетки и УЗИ брюшной полости. Также велись эксперименты по анализу метеосводок.

Упомянутые выше проекты объединяет предложенный Александром Семёновичем семантически-ориентированный подход к анализу запроса и сообщения на ЕЯ, основная идея которого состоит в преимущественном использовании при анализе входного текста лексико-семантической информации (с помощью механизма семантических ориентаций) и структурно-текстовых ограничений (принадлежность к сегменту текста, линейный порядок), а также в отказе от полного морфологического и синтаксического анализа.

В конце 90-х годов под руководством А.С. Нариньяни была создана оригинальная технология лексического анализа Alex, которая позволяет с помощью

Загорулько Ю.А., Кононенко И.С.

Прометей науки. Слово о А.С. Нариньяни

настраиваемых лексических шаблонов произвольной сложности решать задачи поиска в текстовых массивах фрагментов различной степени структурированности и извлечения информации из найденных текстовых фрагментов. Технология успешно прошла тестирование на массиве газетных объявлений о продаже недвижимости, а затем использована для анализа деловых писем в проекте InDoc.

Проект InDoc выполнялся в 2000–2002 гг. В ходе него была создана система интеллектуализации документооборота для крупной инвестиционной компании. Системой осуществлялся тематический анализ и классификация потока входной корреспонденции. В ней семантически-ориентированный подход был реализован в чистом виде, т.е. без учёта морфологии и синтаксиса входного текста, но с существенной опорой на базу знаний об организации, её структуре и деятельности.



Заметим, что извлечение информации из больших массивов документов в ограниченной предметной области является одной из наиболее востребованных задач в широком диапазоне приложений от интеллектуальных справочных и поисковых систем до «продвинутых» составляющих документооборота.

В 2000-е годы А.С. Нариньяни также руководил проводимыми в Новосибирском филиале Российского института искусственного интеллекта исследованиями по разработке технологии КЛАН, предназначенной для создания предметно-ориентированных словарей и их применения при решении различных задач анализа текста в информационных системах. Эта технология предоставляет средства для создания, наполнения и использования словарей, ориентированных на конкретную предметную область. Словари, создаваемые с её помощью, могут поддерживать основные этапы анализа текста (морфологический, синтаксический и семантический), а также осуществлять классификацию потока входных сообщений на основе накапливаемой статистической информации.

В 2001–2002 гг. А.С. Нариньяни опубликовал ряд статей на стыке лингвистики, инженерии знаний и философии. Широкий отклик в научной среде получила серия статей о соотношении онтологии и тезауруса [4, 5], в которой делается вывод о том, что ядром интеллектуальных систем с активным использованием компьютерной лингвистики становится тандем «Тезаурус плюс Онтология».

Завершая первую часть этой статьи, хочется отметить, что А.С. Нариньяни внёс неоценимый вклад в развитие компьютерной лингвистики в нашей стране. Этот вклад заключается не только в его теоретических работах и в прикладных системах, созданных при его участии, но и в формировании сообщества специалистов, объединённых тематикой «компьютерная лингвистика». По его инициативе в 1995 году был возрождён семинар «Диалог», переросший затем в международную конференцию, ставшую главным форумом российских лингвистов. А.С. Нариньяни был душой «Диалога». Ни одно значимое заседание, формальное или неформальное, ни один круглый стол не обходился без его активного участия.

Другой гранью научного творчества А.С. Нариньяни были вопросы представления знаний и их использования в прикладных системах. Наряду с развитием традиционных моделей представления знаний, он внёс большой вклад в разработку принципиально новых методов, базирующихся на учёте так называемых НЕ-факторов. Он заметил, что



«разрабатывая системы, способные адекватным образом воспринимать и использовать данные о всё более сложных фрагментах действительности, мы вынуждены развивать формальный аппарат представления знаний и модели усвоения информации, приближающиеся по своим возможностям к процессу понимания у человека. По мере продвижения в этом направлении становится всё более очевидным, что информация, с которой должны оперировать наши модели, совершенно не соответствует по своему характеру тем формальным объектам, с которыми имеет дело традиционная математика. Если для последних считаются обязательными определённая точность, полнота, замкнутость, непротиворечивость и т.д., то новые — антропоморфные — компоненты аппарата представления знаний отражают те свойства человеческой модели мира, которые характеризуются неполнотой, отсутствием точности, незамкнутостью, возможностью противоречий и т.д.». Поэтому, для того чтобы аппарат представления знаний мог более адекватно отражать свойства моделируемой действительности, он должен как можно более полно учитывать присущие ей неполноту, неточность, неоднозначность и нечёткость наших знаний о ней. Для этого А.С. Нариньяни предложил ввести в модель понимания новые компоненты, которые он назвал НЕ-факторами. Эти совершенно различные по своей природе комплексы свойств он обозначил одним термином «НЕ-факторы» по той причине, что каждый из них отражает компонент модели понимания, являющийся содержательно, а часто и лексически отрицанием одного из классических свойств формальных систем. А.С. Нариньяни даже предположил, что НЕ-факторы образуют нечто вроде периодической системы элементов будущего формального аппарата реальных знаний.

А.С. Нариньяни занимался исследованием многих НЕ-факторов, но наибольший вклад он внёс в разработку «неопределённости» [6], которая играет ключевую роль не только в структуре реальных знаний, но и в приложениях, относящихся к области вычислительной математики. Неопределённость присуща многим нашим знаниям; так, в общем случае доступная информация о фрагменте действительности является неполной даже в рамках фиксированной системы общих знаний. Эта неполнота может касаться как значений величин, так и типов объектов и отношений между ними.

Основываясь на идее практического применения неопределённости, А.С. Нариньяни в начале 80-х годов предложил новый математический аппарат — недоопределённые модели (Н-модели), которые позволяют решать задачи с недоопределёнными (приблизительно заданными) значениями параметров. Рассматриваемый вначале как оригинальный подход в области искусственного интеллекта, этот математический аппарат трансформировался постепенно в прикладную технологию, относящуюся к направлению «программирование в ограничениях» (constraint programming), активно развиваемому в мире в последнее десятилетие в качестве одного из наиболее перспективных в области ИТ. На базе этого аппарата была создана технология Н-моделей, которая выделяется среди других подходов вычислительной мощностью, универсальностью и эффективностью. Фактически она является единственной технологией, которая позволяет решать задачу удовлетворения ограничений в самой общей постановке, что даёт возможность применять её к самым различным классам задач в таких областях, как экономика и финансы, планирование, проектирование и др. На основе этого подхода были созданы прикладные системы UniCalc, Time-Ex, NeMo+ и др., которые широко использовались для решения практических задач.

А.С. Нариньяни был не просто учёным и популяризатором науки, он был ещё и большим подвижником, много сил тратил на то, чтобы достижения нау-



ки находили применение в повседневной жизни. Именно поэтому он выступил одним из инициаторов «Старта» — проекта, направленного на создание отечественных ЭВМ нового поколения (1986–1988 гг.). Он детально изучил аналогичный японский проект создания ЭВМ 5-го поколения, увидел все его достоинства и недостатки. Последние относились большей частью к принципам создания программного обеспечения, где недооценивалась сложность анализа текстов на естественном языке и переоценивалась роль языка логического программирования ПРОЛОГ, вокруг которого строилась и архитектура новых ЭВМ, и их программное обеспечение.



А.С.Нариньяни стал одним из разработчиков программы работ «Старта» по интеллектуальному программному обеспечению ЭВМ нового поколения. Здесь ставилась задача создания не отдельных программных продуктов, а образцов промышленной технологии разработки программных систем, ускоряющих процесс создания таких систем в десятки раз. Стоит отметить, что к концу проекта «Старт» образцы технологий были созданы, но наступившие после этого перемены в стране не позволили довести эти технологии до промышленного использования.

Чтобы продолжать работы по созданию интеллектуального ПО, начатые в рамках «Старта», А.С. Нариньяни в конце 1988 г. создаёт сначала государственную научную фирму «Интеллектуальная технология», а затем, в июле 1991 г., на базе этой фирмы — Российский НИИ искусственного интеллекта с центром в Москве и филиалом в Новосибирске. В рамках этих организаций были продолжены исследования в области представления и обработки знаний и автоматической обработки информации на естественном языке, разрабатывались новые методы решения вычислительных и логико-комбинаторных задач на базе недоопределённых моделей и создавались перспективные программные технологии и прикладные системы.

Большое внимание А.С. Нариньяни уделял применению разработанных при его непосредственном участии Н-технологий для решения практических задач. Вначале в качестве таких задач выступали задачи из области проектирования технических систем, но постепенно фокус сместился на финансово-экономические задачи. Первое время такие задачи решались с помощью универсальной системы программирования в ограничениях UniCalc, а затем была разработана специализированная система ФИМПЛАН — система электронных таблиц, поддерживающая работу с интервальными значениями, своего рода Excel со встроенным в него UniCalc'ом. Только в отличие от системы Excel в электронной таблице Финплана связи между ячейками задавались в виде произвольных алгебраических формул (ограничений), образуя единую математическую модель. Эта модель может иметь как точные, так и частично известные (интервальные) параметры, допускает существование любых зависимостей между параметрами, позволяет совместно использовать линейные и нелинейные уравнения, неравенства и логические выражения. В 1996 г. эта система была с успехом представлена на крупнейшей европейской компьютерной выставке CeBIT (Германия), а в 1997 г. демонстрировалась на международной конференции РАСТ'97 в Лондоне. Позже ФИМПЛАН перерос в технологию INTEGRA.NM, на базе которой решалось множество финансово-экономических задач. В

частности, эта технология была успешно опробована в рамках нескольких проектов Минобороны РФ, при разработке моделей промышленности Москвы, Томской и Ивановской областей, а также экспериментальных моделей экономики республик Казахстан и Болгария.

А.С. Нариньяни был не только одним из организаторов разработки и внедрения информационных технологий — он одним из первых задумался об их влиянии на личность человека, о том, как их повсеместное применение скажется на культуре и поведении будущих поколений, не грозит ли ему новое порабощение на новом, более высокотехнологичном уровне. Этой теме он посвятил целый цикл статей (см. например [7]).

В заключение необходимо сказать, что личность А.С. Нариньяни имеет такой масштаб, что в рамках журнальной статьи можно осветить только какие-то общезначимые, лежащие на поверхности черты. Мы надеемся, что исследование феномена личности А.С. Нариньяни, а также его научного и культурного наследия ещё станет предметом большого научного труда.

Литература

1. Нариньяни А.С. Формальная модель: общая схема и выбор адекватных средств. Препринт ВЦ СО АН СССР, вып. 107. Новосибирск, 1978.
2. Моделирование языковой деятельности в интеллектуальных системах. Сб. научных трудов под ред. А.Е. Кибрика и А.С. Нариньяни. М: Наука, 1987.
3. Нариньяни А.С. Лингвистические процессоры ЗАПСИБ (1-я часть). Препринт ВЦ СО АН СССР, № 199. Новосибирск, 1979.
4. Нариньяни А.С. Кентавр по имени ТЕОН: тезаурус+онтология // Международный семинар по компьютерной лингвистике и её приложениям. «Диалог'2001». 2001. Т. 1. С. 184–188.
5. Нариньяни А.С. ТЕОН-2: от Тезауруса к Онтологии и обратно // Международный семинар по компьютерной лингвистике и её приложениям. «Диалог'2002». Протвино, 2002. Т. 1. С. 307–313.
6. Нариньяни А.С. Неопределённость в системах представления и обработки знаний // Известия АН СССР. Техническая кибернетика. 1986. № 5. С. 3–28.
7. Нариньяни А.С. НОМО — два в одном // Труды международной конференции Диалог'2005 «Компьютерная лингвистика и интеллектуальные технологии». М.: Наука, 2004. С. 457–464.

Сведения об авторах

Загорулько Ю.А. —

кандидат технических наук, заведующий лабораторией Института систем информатики имени А.П. Ершова СО РАН (ИСИ СО РАН).

Кононенко И.С. —

научный сотрудник Института систем информатики имени А.П. Ершова СО РАН (ИСИ СО РАН).