



Интервью с профессором Г. Г. Белоноговым

Искусственный интеллект машинного перевода

В 1996 году Кембриджский биографический центр назвал скромного российского профессора Г. Г. Белоногова человеком года. Как сказано, за выдающийся вклад в мировую компьютерную науку в области машинного перевода. Наш соотечественник просто нашел новый вариант решения этой застарелой проблемы.

Вместо зашедшего в некоторый тупик метода пословного перевода российский профессор предложил принципиально иной метод — фразеологический. Текст в этом случае переводится не отдельными словами, а целыми словосочетаниями. Что в перспективе сулит реально, подчеркиваю, реально адекватный оригиналу перевод. То есть мы на выходе будем получать вместо бессмысленной, как в современных компьютерных переводчиках, абракадабры связный по смыслу и грамматически текст, причем за считанные минуты или секунды.

Но насколько эта заманчивая перспектива осуществима? И как быстро, самое главное? Ответить на эти вопросы, как нам кажется, может только сам создатель теории Герольд Георгиевич Белоногов.

Герольд Георгиевич, в свое время вы предложили и обосновали новый, так называемый фразеологический подход к машинному переводу. В чем его особенность и отличие от традиционного пословного?

Работы по машинному переводу начались с середины 1950-х годов. Машины в то время были очень слабые, поэтому пользоваться можно было только ограниченным набором лингвистических средств, и, естественно, применялся исключительно пословный перевод. Использовались и словосочетания, но преимущественно в идиоматических выражениях. По мере возрастания мощности вычислительной техники подходы слегка модифицировались. Однако первые три десятилетия основной единицей языка, которая хранилась в памяти машины, остава-

» лось слово. Если отдельных слов было заложено в ЭВМ десятки тысяч, то словосочетаний только сотни.

Между тем, в естественных языках слово на самом деле не является основной смысловой единицей, как многие века утверждалось. Основной единицей смысла является понятие. Понятий очень много: по нашим предположениям, в естественном языке их сотни миллионов. Тогда как разных слов всего около одного миллиона. Поэтому большинство понятий выражаются словосочетаниями, причем смысл этих сочетаний, как правило, несводим или не полностью сводим к сумме составляющих их слов. Вот почему, с нашей точки зрения, основными единицами при переводе должны быть не отдельные слова, а словосочетания, выражающие понятия, наименования понятий, отношения между понятиями и фразеологические элементы, отражающие типовые ситуации.

Впервые в печати мысль о том, что перевод должен быть семантико-синтаксическим, преимущественно фразеологическим, была мною сформулирована в предисловии к книге Д. Жукова «Мы — переводчики» еще в 1975 году.

А вот Запад задержался на уровне словословного перевода. Было проведено два эксперимента в Англии: первый Кренфильдский и второй.

А что это были за эксперименты?

В ходе этих экспериментов сравнивалась поисковая эффективность по отдельным словам и словосочетаниям. Они показали, что словосочетания ничего не дают по сравнению со словами. В результате традиционное многовековое заблуждение о слове как основной языковой единице сохраняется во многих странах до сих пор.

Как реализована ваша теория в современных компьютерных переводчиках, прежде всего в системе Retrans Vista? В полной ли мере нашли отражение ваши теоретические изыскания на практике?

В науке, да и в жизни, так устроено, что если идея носится в воздухе, то стихийно этим начинают заниматься разные люди. Раз объективно большинство понятий выражаются словосочетаниями, то даже помимо желаний и утверждений всяких горе-теоретиков фразеологический метод начал применяться на практике.

Вообще, пионерами в области машинного перевода у нас в стране были группы Ю. А. Моторина и Р. Г. Пиотровского. Но еще примерно пять-шесть лет назад у них словарь был всего лишь на 17 тыс. слов и, может быть, несколько тысяч словосочетаний. И только в последнее время они стали использовать готовые словари, которые как раз построены на основе толкования понятий. Поэтому стихийно (они, может быть, сами этого не понимают) у них в значительной степени теперь получается фразеологический перевод.

В нашей же научной группе ВИНТИ (Всероссийского института научной и технической информации) была с самого начала установка именно на такой вариант перевода. Вот почему у нас на сегодняшний день англо-русский и русско-английский словари имеют объем в одном направлении порядка миллиона семисот тысяч словарных статей. Словарь, встроенный в систему Retrans Vista, в англо-русской паре языков самый большой в мире. При этом на перевод однословных единиц в нем имеется всего около 300 тыс., а на все остальные словосочетания от двух слов до семнадцати. Вообще же, по нашим предположениям, в естественных языках наименований понятий сотни миллионов в различных областях человеческой деятельности. Для нормального перевода текста достаточно иметь хотя бы десятки миллионов подобных конструкций.

Многие читатели, наверное, в эту секунду подумают: «А зачем вообще нужен еще один компьютерный переводчик, если он по качеству пока что существенно от того же Promt или «Сократа» не отличается?» В чем преимущество вашей системы по сравнению с конкурирующими продуктами?

Все эти системы испытывались на небольших объемах и в основном либо на бытовой, либо на общественно-политической лексике. Тогда как изначально компьютерные переводчики были ориентированы на научно-техническую информацию. Так вот, эксперименты показывали на простейших примерах (если судить не о грамматике, а о том, как передаются понятия), что при переводе ими либо допускались грубые искажения, либо совсем ничего не переводилось.

Если брать тексты разных тематик, то в среднем неопознанных слов у Promt будет в »



Биографическая справка

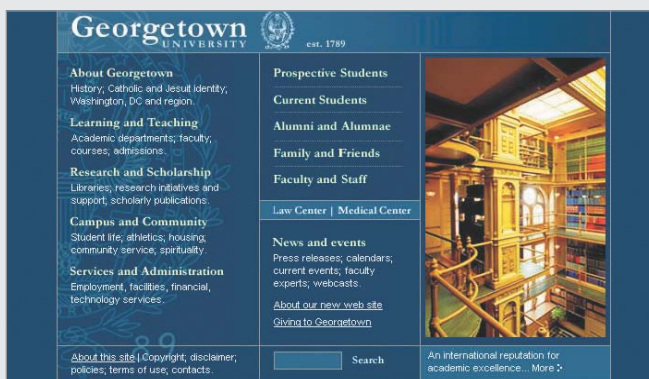
Вехи пути

Герольд Георгиевич Белоногов родился 5 июня 1925 года в селе Солдатово Быстроистокского района Алтайского края. В 1947 году он окончил Военный институт иностранных языков, а затем — Военно-инженерную академию им. В. В. Куйбышева по специальности инженер-электрик. После этого долгие годы работал в одном из военных научно-исследовательских институтов. С 1980 по 2001 год возглавлял Отдел лингвистических исследований во Всероссийском институте научной и технической информации (ВИНИТИ). В настоящее время является научным руководителем и консультантом фирмы «Виста Текнолоджи».

В 1961 году Г. Г. Белоногову была присвоена ученая степень кандидата, а в 1969 — доктора технических наук. В 1995-м он был избран академиком МАИ ИПТ.

В период работы в ВИНТИ Г. Г. Белоноговым были опубликованы десятки научных работ по компьютерной лингвистике. Под его руководством в этом институте было создано несколько систем автоматизированного обнаружения и исправления орфографических ошибок в русских и английских текстах, автоматического индексирования документов и, наконец, фразеологического машинного перевода текстов — с русского на английский язык и с английского на русский. Последняя программа, собственно, и составила основу современной компьютерной системы перевода Retrans Vista, включающей в свой состав самые большие в мире политематические машинные словари. Профессор Г. Г. Белоногов является крупным специалистом в области компьютерной лингвистики и автоматизированных информационных систем. Его вполне можно назвать основателем своей научной школы фразеологического машинного перевода, во всяком случае, им подготовлен уже не один десяток специалистов по этой тематике. Труды Г. Г. Белоногова получили признание во всем мире. Неоднократно он выступал с докладами на научных конференциях в различных странах мира (Россия, США, Франция, Германия, Куба). За выдающиеся научные достижения в 1996 году Кембриджский биографический центр назвал Г. Г. Белоногова человеком года.

Сайт Джорджтаунского университета — места рождения машинного перевода



» пять раз больше, чем у Retrans Vista, а грубых искажений — в три. Мы же отстаем в грамматике, потому что исторически ориентировались прежде всего на русско-английское направление перевода. Но это можно легко подправить буквально за полгода, в отличие от словарей. Там, чтобы улучшить покрытие всего на один процент, надо как минимум удвоить словарь. Как минимум...

Всего-то... Ну тогда понятно, почему, представляя систему Retrans Vista, вы сказали, что идеальный машинный перевод невозможен даже теоретически, что он никогда не сможет достичь уровня профессионального лингвиста. Но тогда получается, что разработчики систем компьютерного перевода работают над тем, чего не может быть...

Это, конечно, очень остроумный вопрос, но потребности-то в переводе научно-технической документации сейчас настолько огромны, что никакой, даже самый квалифицированный переводчик не в силах совладать с этой работой в тех масштабах, которые требуются. Поэтому машинный пе-

ревод, даже не слишком качественный, может стать прекрасным средством повышения уровня информированности научной и деловой общественности о достижениях научных, экономических и т. д. И что еще важнее, если к системе машинного перевода правильно подойти, то она позволит тому же квалифицированному переводчику сэкономить усилия как минимум в два-три раза. То есть перевод будет сделан быстрее и, соответственно, дешевле. И кроме того, как это ни странно, еще и качественнее, потому что реальные тексты всегда будут политематическими. Поверьте переводчику с многолетним стажем, не родился еще такой человек, который бы знал терминологию во всех областях науки и техники. В нашей же системе огромный словарь, обобщающий труд многих тысяч экспертов.

Я слышал, что вы были категорически против разработок в области создания искусственного интеллекта. Почему? Разве машинный перевод не является одним из элементов такого машинного разума?

Это неправильно. Я никогда не был против

создания искусственного интеллекта. Более того, я был членом ученого совета Академии наук по этому вопросу, участвовал в работе таких секций, как «Общение с ЭВМ» и «Компьютерная лингвистика».

Никогда я не отрицал, что многие виды интеллектуальной деятельности человека машина может выполнять лучше и быстрее. Это бесспорно, но часто, когда говорят об искусственном интеллекте, подразумевают вообще замену человека машиной. Так вот эта задача теоретически недостижима. На мой взгляд, машина никогда не заменит нас целиком, ведь она всегда будет лишь нашей моделью. А любая модель всегда беднее оригинала.

Даже один из ведущих отечественных ученых в этой области Г. С. Поспелов предпочитал в последнее время употреблять не термин «искусственный интеллект», а именно «перспективные информационные технологии». Я бы с этим определением согласился.

Так что же, мы никогда не будем общаться с машинами на естественном языке? Как в американских фантастических триллерах, где в космических кораблях обязательно есть какой-то машинный сверхразум, который астронавты ласково называют «дорогушей» и с которым ведут непринужденную беседу. Неужели это так и останется утопией?

Это и утопия, и реальность одновременно. В каких-то предметных областях, которые хорошо изучены и для которых реально смоделировать правильное функционирование объекта, можно будет со временем создать систему диалогового общения с машиной на »



Теория машинного перевода

Имена и проекты

Через два года мы будем отмечать неофициальный юбилей — 50 лет машинному переводу. В далеком 1954 году в Джорджтаунском университете (США) были проведены первые эксперименты, доказавшие тогда саму возможность с помощью ЭВМ переводить тексты с одного языка на другой. К сожалению, возможность эта до сих пор остается нереализованной. И все-таки определенный прогресс достигнут был. Например, сейчас более-менее сносным получается у компьютера перевод деловых и технических бумаг.

У нас в стране пионерами теории машинного перевода были Д. Ю. Панов, А. А. Ляпунов и И. С. Мухин. В их трудах был заложен фундамент, на основе которого в наши дни развиваются такие известные системы компьютерного перевода, как Promt, «Сократ» и недавно появившаяся Retrans Vista. Серьезные исследования в области машинного перевода велись и за рубежом силами таких известных компьютерных лингвистов, как А. Эттингер (США), И. Бар-Хиллер (США) Н. Хомски (Израиль). Наибольшую извест-

ность приобрели труды японского ученого М. Нагао. В свое время он предложил использовать при машинном переводе уже готовые, заведенные в память ЭВМ варианты прочтения исходных текстов, ранее сделанные опытными лингвистами. М. Нагао назвал свой подход Example based translation (перевод, основанный на примерах). Подобная методика, называемая также Translation Memory («Память переводчика»), ныне широко используется при переводе документации в крупнейших корпорациях мира.

» естественном языке. Там будут происходить какие-то квазимыслительные процессы. Все это вполне возможно.

Но все равно это не будет человек. Человек — это явление другого порядка. Он не машина и никогда ею не будет. У нас ведь есть не только мышление, но и интуиция, не только логическое что-то, но и алогическое, не только рациональное, но и иррациональное, не только реальное, но и трансцендентное.

В Советском Союзе, вы об этом только что упомянули, велись научные изыскания в области искусственного интеллекта. Что это был за проект? Какое место отводилось в нем машинному переводу?

В начале 1960-х годов в стране была настоящая эйфория относительно возможностей электронно-вычислительных машин. В принципе, такое в истории часто случалось и раньше. Например, когда изобрели паровоз, то сразу возникла концепция человека-машины. Изобрели компьютер — стали сравнивать человека с ним, хотя это величайшее, по-моему, заблуждение.

Так вот, в качестве типовых задач, которые относили тогда к классу искусственного интеллекта, были поставлены перед учеными следующие: во-первых, автоматическое доказательство теорем, во-вторых, распознавание звуковых и зрительных образов, в-третьих, игра в шахматы и, наконец, в-четвертых, как раз предмет нашего разговора — машинный перевод.

Это тоже, безусловно, интеллектуальная задача. Ведь переводчик сначала читает текст на одном языке, осмысливает его и передает средствами уже другого языка. К



◀ Страничка Всероссийского института научной и технической информации

слову, Британская энциклопедия относит перевод не к ремеслу, а к искусству.

Скажите, если заглянуть лет на двадцать-тридцать вперед, то насколько лучше и качественнее станут, по вашему мнению, системы машинного перевода, та же Retrans Vista? Сможет ли когда-нибудь компьютер сносно перевести Толстого на английский, а Диккенса на русский? Или все это так и останется несбыточной мечтой?

Если бы были созданы минимальные условия для работы нашего научного коллектива, то не через двадцать-тридцать лет, а максимум через два-три года научно-технический перевод в лингвистическом плане был бы на очень высоком уровне. Но это, повторю, может быть только семантико-синтаксический, преимущественно фразеологический перевод.

Что касается художественных текстов, то это моя давняя мечта — попробовать начать переводить их на основе концепции фразеологического перевода. И не исключено, что когда-нибудь за переводчиком ос-

танется только функция редактирования и правки. Даже при переводе поэзии, хотя здесь без последующей обработки, конечно же, обойтись нельзя.

Я не прорицатель, но через два-три года, ну от силы пять лет, как мне кажется, компьютерные переводчики будут использоваться в качестве основного средства, помогающего нам при работе с художественной литературой на иностранных языках. Но высшим судьей все равно будет человек.

«Через четыре года здесь будет городсад», — помнится, в свое время обещал нам поэт Маяковский. Профессор Белоногов через пять лет обещает представить адекватный машинный перевод с одного языка на другой. Если, конечно, будет соответствующее финансирование проекта. Ну что ж, проживем, как говорится, увидим. И немного помечтаем о мире без языковых проблем. Может быть, более спокойном мире, мире всеобщего взаимопонимания. В прямом смысле слова или словосочетания.

■ ■ ■ Беседовал Александр Евдокимов

ПОДПИСКА 2002

1 Объединенный каталог «Пресса России», индекс 44077

2 Каталог «Газеты, журналы» Агентства «Роспечать», индекс 80978

Кто хочет знать, читает СНГР