

УДК 519.7

Н.А. ГВОЗДИНСКАЯ, З.В. ДУДАРЬ, Ю.П. ШАБАНОВ-КУШНАРЕНКО

О МАТЕМАТИЧЕСКОМ ОПИСАНИИ СМЫСЛА ТЕКСТОВ ЕСТЕСТВЕННОГО ЯЗЫКА

С изобретением паровой машины человечество в XVII столетии вступило в эпоху *первой научно-технической революции*, которая характеризуется интенсивным изучением физических процессов и созданием технических средств для их искусственного воспроизведения. С изобретением ЭВМ в середине текущего столетия началась *вторая научно-техническая революция* (первая же не прекратилась, она продолжается), которая характеризуется интенсивным изучением информационных процессов и созданием технических средств для их искусственного воспроизведения. Результатом первой научно-технической революции стала механизация физического труда людей, результатом второй — механизация умственного труда. Технические системы, механизующие умственный труд людей, получили название *систем искусственного интеллекта*. Они стали главным средством искусственного воспроизведения информационных процессов.

Термин “искусственный интеллект” не следует понимать слишком буквально: современные технические системы, обрабатывающие информацию, во многих отношениях уступают по своим интеллектуальным возможностям разуму человека. Важнейшим недостатком искусственного интеллекта, существенно ограничивающим сферу его практического использования, является неспособность машины понимать человеческую речь и, как следствие, невозможность смысловой обработки ею текстов естественного языка. Хотя все люди свободно владеют естественным языком, однако они не знают, как им это удастся, и потому не могут передать машине эту свою важнейшую интеллектуальную функцию. Успешное изучение механизма языка и его адекватное математическое описание, в особенности разработка эффективного метода формального описания смысла (значения) текстов, создали бы научную базу для освоения машиной естественного языка, и, в конечном итоге, привели бы к коренному усовершенствованию систем искусственного интеллекта.

Считается, что процесс понимания текстов естественного языка пока не поддается формализации. Язык характеризуется как “колдовское действо” [1, с.33]. Для эффективного проникновения в механизм языка нужен какой-то принципиально новый подход к его изучению, который видится в сравнении естественного языка с языком математики. Основанием к этому

служит предположение о существовании глубокой аналогии, близкого родства между этими двумя языками. Математический язык, в отличие от естественного, хорошо формализован, его механизм понятен. Он создавался не стихийно сам по себе, а целенаправленно отдельными людьми и поэтому весь на виду. Если гипотеза о наличии глубокой аналогии между этими двумя языками оправдается, тогда можно будет найти ответы на многие трудные вопросы, касающиеся механизма естественного языка, обращаясь за подсказкой к языку математики.

Приступая к изучению механизма языка, часто начинают с вопроса: "Что такое смысл предложения?". И вопрос этот так и остается без ответа. Здесь мы попытаемся на него ответить, обращаясь за подсказкой к языку математики. Математические тексты, как и тексты естественного языка, состоят из предложений. Предложения, фигурирующие в математических текстах, называют высказываниями. Зададимся вопросом: "Что такое смысл высказывания?" Ответить на него нетрудно: как известно, *смыслом высказывания* является функция зависимости истинностной переменной от предметных переменных этого высказывания. Любое высказывание бывает либо истинным или ложным. Если высказывание истинно, то ему приписывают значение 1, называемое *истиной*; если ложно — то значение 0, называемое *ложью*. Образует множество $\Sigma = \{0, 1\}$, его элементы называются *логическими*. *Истинностной* называется переменная ξ , определенная на множестве Σ . Говорят, что переменная $\xi \in \Sigma$ принимает *истинностные значения* 0 и 1. В роли *предметных* могут выступать любые переменные, которые используются в математике и встречаются в ее текстах.

Рассмотрим, например, высказывание " $x+y=z$ " (а), где x, y, z — предметные переменные, под которыми в данном случае понимаются переменные, определенные на множестве $N = \{0, 1, 2, \dots\}$ всех неотрицательных целых чисел. В таком виде высказывание (а) ни истинно, ни ложно. Но если подставить в него вместо переменных x, y, z какие-нибудь их конкретные значения из множества N , то высказывание (а) обратится в истину или ложь. Пусть, к примеру, $x=2, y=3, z=5$. Тогда высказывание (а) обращается в *тавтологию* $2+3=5$, и говорят, что оно *истинно*. Если же принять $x=0, y=1, z=2$, то это же высказывание обратится в *противоречие* $0+1=2$, в этом случае говорят, что оно *ложно*. Мы видим, что высказывание (а) включает в себе некоторую функцию $\xi = f(x, y, z)$, отображающую множество N^3 в множество Σ . Функции с двоичными значениями, зависящие от произвольных аргументов, называются *функциями истинности* или *предикатами*. Итак, смыслом математического предложения (высказывания) является предикат $\xi = f(x_1, x_2, \dots, x_m)$, представляющий собою зависимость истинно-

стной переменной ξ от предметных переменных x_1, x_2, \dots, x_m этого высказывания. Любое математическое утверждение представляет собой уравнение вида $f(x_1, x_2, \dots, x_m)=1$, которое задает вполне определенное отношение, связывающее между собою предметные переменные x_1, x_2, \dots, x_m [2]. Важно отметить, что такая трактовка смысла высказывания в математике общепринята, никаких сомнений в ее правильности ни у кого не возникает. Математик рассматривает каждое высказывание как формулу, а любая формула, в чем он твердо убежден, всегда выражает какую-то функцию и ничего более. Именно для выражения функций и были придуманы формулы математиками. Высказывание же всегда выражает некоторую функцию с двоичными значениями, таким образом его смыслом (значением) служит вполне определенный предикат.

Мы рассмотрели, что представляет собой смысл математического предложения. Теперь, руководствуясь гипотезой о близком родстве естественного и математического языков, предположим, что такая трактовка смысла применима и к предложениям естественного языка (мы имеем в виду только повествовательные предложения). Иными словами, мы исходим из того, что смыслом каждого предложения естественного языка служит некоторый предикат. А если это так, то к каждому предложению естественного языка должно быть применимо понятие истинностной переменной с ее двумя значениями “истинно” и “ложно”. Проверяем этот вывод на практике. Берем какое-нибудь предложение естественного языка, например “На столе лежит книга”. Взятое само по себе, оно ни истинно, ни ложно. Но если его соотнести с какой-нибудь реальной ситуацией, то предложение, и в самом деле, становится истинным или ложным.

Справедливость последнего утверждения продемонстрируем следующим экспериментом. Исследователь, обращаясь к испытуемому, произносит предложение: “На столе лежит книга”. В роли испытуемого может выступать любой человек, владеющий русским языком. В тот момент, когда исследователь произносит слово “стол”, он указывает на некоторый предмет (назовем его первым); когда же проговаривает слово “книга”, указывает на второй предмет. Предметы выбираются исследователем произвольно: они не обязательно должны быть столом и книгой. Исследователь просит испытуемого определить, соответствует содержание предложения фактическому положению дела или нет; иначе говоря, он предлагает ему установить, истинно или ложно в данной конкретной ситуации предъявленное предложение. Ясно, что если первый предмет есть стол, а второй — книга, и второй, к тому же, лежит на первом, то испытуемый даст положительный ответ; если же хотя бы одно из этих условий не выполняется (например, второй предмет — это тетрадь, или же — книга, но не лежащая на столе, а

стоящая в шкафу), тогда испытуемый ответит отрицательно. Итак, мы видим, что в данном вопросе гипотеза о близком родстве естественного и математического языков себя оправдывает: оказывается, что понятие истинностной переменной к предложениям естественного языка применимо, подобно тому как оно применимо к предложениям языка математического.

Движемся дальше. Если смыслом предложения естественного языка, действительно, служит предикат, то в предложениях должны присутствовать не только истинностные, но также и предметные переменные. Проверим, так ли это на самом деле. Берем предложение “На столе лежит книга”. Ищем в нем предметные переменные. Но их там нет! Что же делать? Отказаться от принятой нами гипотезы, а значит и от задачи математического описания смысла текстов естественного языка? Не будем торопиться с этим. Воспользуемся известным принципом: “Факты не подтвердили нашу гипотезу? Тем хуже для фактов!” Разберемся более тщательно, о чем говорят факты. Они свидетельствуют лишь о том, что предметных переменных нет в *тексте* предложения. Но текст предложения — это еще не само предложение, это — всего лишь его запись. Предложение характеризуется не только текстом, но и своим *смыслом*, оно выражает некоторую мысль. А мысль находится в уме человека. Есть ли гарантия того, что все детали смысловой стороны предложения выражены его текстом? Такой гарантии, конечно же, нет. Известно, что в текстах естественного языка очень широко используются сокращения (к ним относятся, например, эллипсис, анафора и омография). Математические выражения тоже пишутся в сокращенной форме [2]. Не исключено, что в тексте предложения предметные переменные просто, для краткости, опущены. А в мысли, выражаемой предложением, предметные переменные, быть может, и отыщутся.

Поискем предметные переменные в мысли, соответствующей предложению “На столе лежит книга”. Достаточно нацелить себя на решение такой задачи, и она сразу же решается: предметные переменные тот час же обнаруживаются сами собой. Фактически ими мы уже оперировали, когда рассматривали эксперимент на испытуемом. В нем исследователь указывал на два предмета, причем каждый из них он выбирал произвольно, т.е., на самом деле, там речь шла не только о предметах, но и о предметных переменных. Обозначим последние символами x и y . Когда исследователь указывает испытуемому конкретные предметы (например, какой-то стол и какую-то тетрадь), он приписывает переменным x и y определенные значения. Можно ли записать предложение в более развернутой форме так, чтобы в нем появились предметные переменные? Да, можно. И это совсем нетрудно сделать. Выглядеть теперь оно будет следующим образом: “Предмет x есть стол, предмет y есть книга, предмет y лежит на предмете

х". Очевидно, что в такой, более подробной, записи предложение выражает ту же самую мысль, что и его краткая исходная запись. Ясно, что каждый человек, устанавливающий, истинно или ложно некоторое предложение, для решения этой задачи будет вынужден обратиться к предметным переменным этого предложения и заместить их конкретными предметами из некоторой вполне определенной ситуации. Текст предложения естественного языка мы можем рассматривать как формулу (записанную, как мы только что убедились, в сокращенной форме), которая выражает некоторый предикат. По ней человек вычисляет истинностное значение предложения при заданных значениях его предметных переменных.

Итак, мы видим, что гипотеза о тесном родстве естественного и математического языков снова себя оправдала. С ее помощью мы обнаружили в предложении естественного языка не только истинностную переменную (о чем науке уже давно было известно), но также и не замеченные никем (если полагаться на выполненный нами анализ литературных источников) до сих пор предметные переменные. В результате подтвердилось предположение, что предложение естественного языка является, как и математическое высказывание, носителем некоторого предиката, представляющего собой зависимость истинностной переменной предложения от его предметных переменных. Этот предикат мы и примем в качестве формального эквивалента смысла предложения естественного языка. Резюмируя все, сказанное выше, мы приходим к выводу, что в непреступной стене, преграждавшей путь машине к освоению ею человеческого языка, теперь образовалась брешь. Заглянув в нее, мы видим, что в понятии смысла предложения естественного языка нет ничего таинственного: смысл каждого предложения представляется нам только предикатом, значения которого определяются значениями его предметных переменных. Мы видим, что предложения естественного языка, как и математические утверждения, в содержательном плане представляют собой отношения — и ничего сверх этого. Оговоримся, однако, что сделанный вывод — это пока не более, чем сугубо предварительное заключение, вынесенное на основании лишь беглого взгляда на открывшуюся новую безбрежную (как нам представляется) область исследований. Если наши предположения оправдаются, то в дальнейшем предстоит планомерное изучение смысла текстов естественного языка с целью его математического описания, направляемое гипотезой о наличии глубокой аналогии между естественным и математическим языками.

Утверждение о том, что смысл любого предложения естественного языка и текста, образованного из предложений, можно исчерпывающим образом формально выразить некоторым вполне определенным предикатом, порождает массу возражений и вопросов. С их рассмотрения мы и

начнем разработку семантической теории естественного языка. Одно из главных возражений состоит в следующем. Если каждое предложение имеет свой единственный смысловой предикат, то как могут существовать многозначные предложения, например предложение “Какой стол, такой и стул”, имеющее два смысловых значения? Ответ, в сущности, очень прост: многозначность предложения проистекает исключительно из-за того, что его текст записан в излишне сокращенной, а потому — и неполной, форме. Возьмем, к примеру, рассматривавшееся нами ранее математическое высказывание “ $x+y=z$ ” и “сократим” его, исключив из его записи знак $+$. В результате получим запись “ $x y=z$ ” с пропущенным знаком операции. Ее можно понимать по-разному: как прежнее высказывание, как “ $xy=z$ ” или “ $x-y=z$ ”, или еще каким-то иным. Математик никогда не согласится принять такую сокращенную запись за само высказывание. Он скажет, что запись высказывания многозначна из-за своей неполноты, поэтому невозможно установить точно, что это за высказывание. Можно лишь говорить о его возможных вариантах. Смысл высказывания можно будет определить однозначно лишь после уточнения его текста.

Представляется, что правильный подход к изучению проблемы многозначности смысла предложения естественного языка состоит в четком разграничении двух понятий: текста предложения и мысли (или мыслей), выражаемой им. Текст предложения может быть многозначным, но каждая из выражаемых им мыслей всегда единична. Иногда случается, что многозначность в текст предложения вводят преднамеренно, как в случае с вышеприведенной фразой. Это значит, что таким предложением хотят выразить одновременно несколько мыслей (в данном случае — две, в одной из них речь идет о мебели, в другой — о питании, из-за чего фраза приобретает юмористический оттенок). Мы обнаруживаем, что каждому предложению соответствует, вообще говоря, не одна мысль, а целое их множество (в частности, если предложение однозначно, то это множество будет одноэлементным, если противоречиво, то — пустым). Будем однозначные предложения и тексты в дальнейшем называть *высказываниями*, заимствуя этот термин из математики (каждое математическое высказывание однозначно). Примером естественно-языкового высказывания может служить использованное нами ранее предложение: “Предмет x есть стол, предмет y есть книга, предмет z лежит на предмете x ”. Предстоит большая работа по созданию методов поиска всех высказываний, охватываемых каждым конкретным предложением, и записи их смысла (единственного для каждого высказывания) на формальном языке.

Возьмем фразу “На столе лежит книга”. Она включает в себе много смысловых вариантов. Один из них мы уже рассмотрели ранее. Рассматри-

ваем другой вариант. Он отличается от первого тем, что теперь исследователь, предъявляя испытуемому ту же фразу, указывает ему только первый предмет. Если этот предмет не стол, то испытуемый отреагирует отрицательным ответом, а если стол, то посмотрит, лежит ли на нем хотя бы одна книга. Если *да*, он скажет, что высказывание истинно, а если *нет*, то — ложно. Полный смысловой перевод фразы теперь получается несколько иным: “Предмет x есть стол, и существует предмет y такой, что предмет y есть книга, и предмет y лежит на предмете x ”. Есть и третий вариант, когда исследователь указывает только на второй предмет. Полный смысл фразы в этом случае запишется в виде: “Предмет y есть книга, и существует предмет x , такой, что предмет x есть стол, и предмет y лежит на предмете x ”. Если же исследователь, предъявляя все ту же фразу, не указывает ни одного из предметов, тогда испытуемый будет вынужден обратиться к четвертому смысловому варианту: “Существуют предметы x и y такие, что предмет x есть стол, предмет y есть книга, и предмет y лежит на предмете x ”. Чтобы определить истинностное значение этого высказывания, испытуемый должен обзреть ситуацию, например, комнату, в которой проводится эксперимент (исследователь еще до начала опытов обязан четко очертить ситуацию, в которой будет действовать испытуемый), выявить все наличные в ней столы, и, если хотя бы на одном из них лежит, по крайней мере, одна книга, отреагировать положительным ответом, в противном случае — отрицательным. Итак, мы видим, что предложение “На столе лежит книга” содержит в себе четыре различных высказывания. Это предложение играет роль *неполной записи* для данных высказываний. Каждое из этих высказываний выражает единственную заключенную в нем мысль и представляет собой ее *полную запись*. Мы видим, что, кроме таких, известных в языкознании, причин многозначности предложения, как эллипсис, анафора и омография, есть и еще одна — отсутствие предметных переменных в тексте предложения.

Незаметно, как бы между прочим, мы научились записывать *естественно-языковые высказывания*, т.е. однозначные в смысловом отношении тексты естественного языка. Этого удалось достичь введением в текст предложения предметных переменных. Нам представляется, что это — существенное достижение. (Вопрос о том, является данное предложение высказыванием или нет, решается с помощью эксперимента на испытуемом. Если, отправляясь от предъявленного ему предложения, испытуемый своим поведением воспроизводит для любой конкретной ситуации единственный предикат, то такое предложение является высказыванием, если нет, то — не является). Но в том виде, в котором тексты высказываний нам удавалось до сих пор записывать, они обладают серьезным недостатком: им

еще очень далеко до полной формализации, поскольку записываются они все на том же естественном языке. Правда, в текстах высказываний появились вкрапления математической символики в виде предметных переменных. Такие тексты понимает человек, но не сможет понять машина (слово “понять” в данном случае означает: выразить смысловой предикат высказывания в виде чисто математической формулы без использования в ней каких бы то ни было естественно-языковых единиц).

Двигаться в направлении полной формализации записи высказывания можно, постепенно заменяя в ней элементы естественного языка на математические выражения. Процесс этот не может продолжаться бесконечно, рано или поздно он должен исчерпаться. И тогда мы получим полностью формализованную запись высказывания, которая будет доступна пониманию машины. Сейчас мы сделаем еще один шаг в этом направлении. Рассмотрим высказывание: “Предмет x есть стол, предмет y есть книга, предмет y лежит на предмете x ”. Оно легко расчленяется на три высказывания: “Предмет x есть стол”, “Предмет y есть книга”, “Предмет y лежит на предмете x ”. Каждое из этих высказываний имеет свой смысл, а значит, — и свой предикат. Обозначим предикат первого высказывания символом $\text{стол}(x)$, второго — $\text{книга}(y)$, третьего — $\text{лежит на}(y, x)$. Записи стол , книга , лежит на теперь выступают в роли имен каких-то предикатов, а каких именно мы еще не знаем. Ясно, что предикат исходного высказывания можно представить в виде конъюнкции этих трех предикатов. Смысл исходного высказывания теперь запишется в виде $\text{стол}(x) \wedge \text{книга}(y) \wedge \text{лежит на}(y, x)$. Хотя до полной формализации текста этого высказывания все еще далеко, тем не менее, некоторое продвижение вперед имеется. Последняя запись уже похожа на формулу: видно, что ею выражается какой-то предикат. Вместе с тем, она легко воспринимается и как само высказывание: по ней можно без труда восстановить исходный текст высказывания. Нелишне заметить, что запись высказывания в результате перехода к словам-предикатам существенно сократилась: теперь она стала лишь немного длиннее первоначальной фразы “На столе лежит книга”. Она, в отличие от последней записи, имеет единственный смысл. Полная формализация данного текста высказывания будет достигнута, если удастся избавиться в нем от словесного материала стол , книга и лежит на , заменив его математическими выражениями.

Но это не предел того, что можно легко сделать уже сейчас для формализации текста. Предикат $\text{лежит на}(y, x)$ можно подвергнуть дальнейшему расчленению. Введем предикат $\text{лежит}(y)$, понимаемый в смысле “Предмет y лежит”, и предикат $\text{на}(y, x)$, соответствующий высказыванию “Предмет y

расположен на предмете x ". Ясно, что предикаты **лежит** $\text{на}(y, x)$ и **лежит** $(y) \wedge \text{на}(y, x)$ совпадают. Допускает дальнейшее расщепление и предикат **лежит** (y) . Его можно представить в виде конъюнкции предикатов **настоящее время** (y) и **лежать** (y) . Первый предикат означает: "Предмет y наблюдается в текущий момент времени", второй — "Предмету y присуще состояние лежания" (безотносительно ко времени). Соединяя конъюнкцией все введенные предикаты, записываем рассматриваемое высказывание в виде: **стол** $(x) \wedge$ **книга** $(y) \wedge$ **настоящее время** $(y) \wedge$ **лежать** $(y) \wedge$ **на** (y, x) . Двигаться дальше в анализе смысловой структуры данного высказывания мы пока не будем. Чтобы это осуществить, надо будет проникнуть в смысловую структуру слов "стол", "книга", "настоящее время", "лежать" и "на". Для этого можно обратиться к толковому словарю, выразив смысл каждого из этих слов с помощью приведенных в нем определений, в роли которых выступают словосочетания, составленные из других слов. Когда этот процесс оборвется, нужно собрать слова, оставшиеся невыраженными, и связать их друг с другом системой высказываний, воспринимаемых носителями естественного языка как истинные. Эта система высказываний выполняет роль аналога системы аксиом, определяющих первичные понятия математики. А в языке она абстрактно определяет смысл слов, которые не удастся выразить с помощью прямых определений через другие слова.

Список литературы: 1. Звезинцев В.А. Предложение и его отношение к языку и речи. — Изд-во Моск. ун-та, 1976. 305с. 2. Бондаренко М.Ф., Дударь З.В., Шабанов-Кушнаренко Ю.П. Отношения как предмет формульного описания // АСУ и приборы автоматки. Вып. 107. С. 94—103.

Поступила в редколлегию 16.12.97