

МОДЕЛИ КОГНИТИВНЫХ ПРОЦЕССОВ

(Вычислительные системы)

1997 год

Выпуск 158

УДК 519.766

О ВВЕДЕНИИ МЕТАУРОВНЯ В ФУНКЦИОНАЛЬНЫХ МОДЕЛЯХ ЕСТЕСТВЕННОГО ЯЗЫКА ¹

М.К. Тимофеева

1. Содержание статьи входит в область функционального моделирования естественного языка, т.е. имитации языка в действии — в процессе диалога. Минимальным диалогом предлагается считать такое взаимодействие двух участников, при котором первый посредством предъявления текста (в устной или графической форме) инициирует выполнение вторым некоторых действий, среди этих действий выделяются умственные действия — их необходимо произвести для того, чтобы "понять" предъявленный текст; действия, являющиеся результатами понимания текста, выходят за рамки языка и здесь не рассматриваются. Произвольный диалог — это последовательность минимальных диалогов. Осмысленность диалога зависит от того, в какой степени он представляет собой именно взаимодействие его участников, насколько предъявитель текста отдает себе отчет в том, какие действия он вынуждает своего собеседника предпринять при понимании этого текста. Для оценки такой осмысленности не достаточно грамматической и семантической правильности построения текста. Предлагается оценивать степень осмысленности диалога в зависимости от способности первого его участника вообразить, а,

¹Работа поддержана РФФИ, грант № 96-06-80970

возможно, и предсказать те умственные действия, в которых может состоять понимание текста вторым участником. (Если такая способность полностью отсутствует, то неясно, в каком смысле данную ситуацию вообще можно назвать диалогом.)

Основная сложность при формализации процесса понимания текстов естественного языка — их неоднозначность: один и тот же текст в разных ситуациях и разными людьми может пониматься по-разному. Чем неоднозначнее текст, тем труднее построить модель, допускающую указанное понимание осмысленности диалога. Если же модель диалога не обладает таким свойством, то она не позволит приписать участнику диалога ту меру ответственности за предъявление именно данного, а не какого-либо другого, текста, которая необходима в конкретной ситуации совершения диалога. (Требуемая мера ответственности может быть достаточно высока, например, в случае модели диалога, предназначенной для управления некоторой сложной системой.)

Имеющиеся модели диалога неявно основываются на предположении о том, что можно построить достаточно надежные формальные средства разрешения неоднозначностей текстов естественного языка, введя ограничения на используемые языковые конструкции и тематику текстов (предметную область). Число предметных областей практически бесконечно, языковые средства, формализацией которых можно заниматься, многочисленны, поэтому велико и число существующих моделей ограниченного таким образом естественного языка. Жесткие ограничения предметной области и языковых средств вообще трудно соблюдать в реальном диалоге, поэтому обычно, в той или иной степени, предусматривается возможность нарушения этих ограничений. Построить практически полезные ориентиры в многообразии существующих моделей языка нереально, так как эти модели, как правило, предназначены для своих узких целей, средства же их формального сопоставления или совмещения отсутствуют. Несмотря на разобщенность попы-

ток формализации естественного языка и многообразие используемых в них формализмов, по своей сложности имеющиеся модели, в подавляющем большинстве, могут быть сведены к логике предикатов первого порядка.

Построение любой модели, как некоторой гипотезы о языке, влечет за собой предположения о том, какими навыками должен владеть носитель языка для того, чтобы вообще быть способным усвоить и использовать язык, удовлетворяющий данной гипотезе. Образование таких навыков или "языковых умений" предшествует усвоению языка, объектом же их применения является сам язык или его модификация, поэтому эти "языковые умения" можно отнести к метаязыку. Существующие модели языка иногда включают в себя части, инвариантные относительно рассматриваемой предметной области, уровень же "языковых умений" в них вообще не анализируется. Попытка формализации хоть каких-то составляющих этого уровня не только представляла бы самостоятельный интерес, но и была бы практически полезна для построения более общих моделей языка.

Данная статья предлагает изменить в этом отношении сложившуюся традицию и обратиться к уровню "языковых умений", аргументируя это тем, что введения этого уровня все равно не избежать, если мы хотим строить модель диалога, удовлетворяющего приведенному требованию осмысленности. Подвергается сомнению предположение о том, что можно выполнить это требование, сузив предметную область и ограничив сложность используемых текстов. Приводится класс неоднозначностей, остающихся в текстах естественного языка при любом подобном его "упрощении", и, следовательно, всегда сохраняющих опасность того, что модель диалога не будет удовлетворять требованию осмысленности. В качестве выхода из данной ситуации и предлагается обратиться к метаязыку языка.

2. Модель естественного языка, в целом, можно рассматривать как способ описания смысла выражения "знать язык x ". Причем, если (как это обычно делается) для каждого ограниченного варианта некоторого естественного языка строится своя модель, то тем самым неявно предполагается, что в этом выражении слово "язык" вообще не значимо. Иначе задавался бы также вопрос о том, что означает выражение "знать язык" независимо от того, о каком языке идет речь. Постановка этого вопроса в его полном объеме сейчас была бы нереальна, поэтому в статье он ставится гораздо более узко: существует ли вообще какая-либо необходимость вводить метауровень при построении функциональных моделей естественного языка, если, в целом, следовать имеющемуся направлению их развития?

Тот класс неоднозначностей текстов естественного языка, который пытаются минимизировать ограничивая предметную область и сложность текстов, включает, в основном, неоднозначности сопоставления тексту формулы, описывающей его смысл. Число таких неоднозначностей действительно может быть уменьшено путем ограничения тематики и сложности текстов. Однако, есть и другой класс неоднозначностей — те, которые возникают уже после того, как тексту были поставлены в соответствие формулы, выражающие его допустимые смыслы, и была выбрана одна из них, — неоднозначности, проистекающие из того, что интерпретацию этих формул можно строить по-разному. Указанное требование к осмысленности диалога может быть предъявлено только в том случае, если предъявитель текста способен вообразить все возможности разрешения неоднозначностей вторым участником диалога. Будем считать, что он обладает такой способностью только в отношении тех неоднозначностей, которые он может каким-либо образом воспроизвести в своем сознании — либо явно определив их в виде выражений используемого им языка, либо изменив свою интерпретацию таких выражений. Предполагается, что указанные способы упрощения естественного

языка позволят эксплицитно выразить каждый вариант понимания текста, содержащего неоднозначности первого класса. Поэтому, если бы имелся только этот класс неоднозначностей, то можно было бы строить модели диалога, удовлетворяющие требованию осмысленности, и без введения какого-либо дополнительного метауровня. Среди же вариантов понимания текста, содержащего неоднозначности второго класса, существуют такие, которые не могут быть эксплицитно выражены на естественном языке и в отношении которых человек не волен изменять имеющуюся у него и интерпретацию. (Это поясняется приводимыми ниже примерами.) Способы разрешения таких неоднозначностей предьявитель текста воспроизвести не сможет. В этом случае осмысленности диалога можно добиться путем введения метауровня. Если исходить из предположения о том, что для построения модели естественного языка следует использовать логику предикатов первого порядка, то обсуждаемый здесь вариант метауровня может рассматриваться как необходимое условие адекватности модели такого типа, т.е. как тот набор навыков, которыми должен обладать носитель языка, описываемого посредством логики первого порядка.

3. Функциональную модель естественного языка, построенную на базе логики предикатов первого порядка, можно изобразить в виде тройки $\langle G, M, I \rangle$, где G — некоторое исчисление предикатов первого порядка, M — предметная область, I — отображение, ставящее в соответствие G его интерпретации на M . Примером такой модели могут служить языки действий [1,2] — класс формальных языков, тексты которых внешне не отличаются от текстов ограниченного (по предметной области и сложности) естественного языка, а G — многосортная логика предикатов первого порядка. Языки действий подробно описаны в [1,2], они используют некоторые принципы, соотносимые с такими направлениями описания естественного языка как категориальные грамматики [3], унификационные грамматики и Gapping Grammars [4]. Язык

действий L , моделирующий ограниченный естественный язык, описывается как функция, которая, в принципе, может преобразовывать любую из частей $L - G, M, I$ (однако, посредством применения языка могут осуществляться не все такие преобразования). Каждая осмысленная часть текста T этого языка (например, слово) определяет стратегию осуществления таких преобразований. В конечном итоге, после анализа всего предъявленного текста, эта стратегия описывается формулой исчисления G . Такая формула-стратегия является недоопределенной, она задает на самом деле только тип преобразования. Дальнейшее доопределение этой формулы до описания конкретного преобразования осуществляется путем интерпретации формулы-стратегии отображением I и последующего применения правил, определенных в M . Предметная область M задается множеством преобразований, описывающих то, как M могло бы развиваться само по себе не будь к нему никаких обращений на естественном языке (т.е. естественный язык рассматривается лишь как одно из внешних средств изменения M).

Допустим, что происходит диалог между s_1 и s_2 (возможно, что $s_1 = s_2$). Пусть s_1 и s_2 пользуются языком, описываемым одним и тем же исчислением G , и говорят об одной и той же предметной области M . Язык L реально будет выступать здесь в двух вариантах: $L_1 = \langle G, M, I_1 \rangle$ и $L_2 = \langle G, M, I_2 \rangle$, первым из которых пользуется s_1 , а вторым — s_2 . Множества текстов, правильных с позиций языков L, L_1, L_2 , совпадают. Каждый текст T языка L в языках L_1 и L_2 рассматривается как одна и та же формула-стратегия f , в ходе же интерпретации эта стратегия дополняется до двух описаний конкретных преобразований: s_1 , в момент предъявления текста T , "понимал" его как выполнение преобразования f_1 в L_1 , s_2 "понял" его как f_2 в L_2 . Само по себе различие преобразований f_1 и f_2 не означало бы неосмысленности данного диалога между s_1 и s_2 , если бы множество конкретных преобразований, соответствующее формуле f в L_1 , содержало f_2 (т.е. для s_1 было бы возможно "вообразить" f_2),

но, если формула f содержит хотя бы одно выражение, которому отображения I_1 и I_2 ставят в соответствие разные элементы M , то это может быть и не так.

Предположим, что языки L, L_1, L_2 обладают свойством композициональности, т.е. семантика любой сложной формулы из G является функцией от семантик входящих в нее подформул. Преобразования f_1 и f_2 получены в результате интерпретации формулы f . Можно построить такие функции g_1 и g_2 , что для любого $A \in G$, входящего в формулу f_1 , существуют такие $B, C \in G$, входящие в формулы f_2 и f , соответственно, что выполняются $g_1(B) = C, g_2(C) = A$, т.е. g_1 ставит в соответствие символу B его образ в f относительно отображения I_1 , g_2 ставит в соответствие символу C его образ в f относительно отображения I_2 . Преобразования f_1 и f_2 функционально изоморфны относительно суперпозиции $g = g_1 \circ g_2$, т.е., если A и B — некоторые подформулы f_1 , причем B содержит A как подформулу, то $g(A), g(B)$ — подформулы f_2 , причем $g(B)$ содержит $g(A)$ как подформулу.

4. Функциональный изоморфизм составляет одну из основных проблем, обсуждающихся в связи с функциональным подходом к моделированию человеческого поведения [5], так как в силу функционального изоморфизма неразличимыми могут оказаться вещи совершенно различной природы. Это создает проблемы и при моделировании естественного языка. Как бы мы ни ограничивали сложность моделируемого фрагмента языка и предметную область, отображения I_1 и I_2 всегда могут оказаться таковы, что "понимания" текстов u_1 и u_2 будут функционально изоморфны и "понимание" текста u_2 будет недоступно для s_1 .

Обычно при построении моделей диалога предполагается, что интерпретации формул G , если и различаются, то настолько мало, что этим можно и пренебречь, а если уж различия между ними сильны, то этот факт не останется незамеченным — он проявит себя тем, что диалог будет выглядеть внешне совершенно бессмысленным. Следующий пример из [6] показывает, что это мо-

жет быть и не так. Некоторые различия интерпретации не могут быть обнаружены ни при какой длине диалога.

В любом фрагменте естественного языка есть слова, именующие наши восприятия: названия цветов, звуков, видов поверхностей и т.д. Рассмотрим, например, цвета. Кажется вполне осмысленным предположить, что могут существовать два человека таких, что объекты, которые они оба называют красными, выглядят для первого из них так же, как объекты, которые они оба называют зелеными, выглядят для другого. Можно предположить, что существуют инвертирующие линзы, которые, если их поместить в глаз человека, приводят к восклицаниям типа "красные предметы теперь выглядят для меня так, как обычно выглядели зеленые, и наоборот". Вообразим теперь, что такие линзы вставлены одному из близнецов при рождении. Близнецы нормально развивались и, повзрослев, стали функционально эквивалентными. То, что их восприятия цветов совершенно различны, не может быть никак обнаружено по их поведению. Такую ситуацию можно назвать инверсией стимулов. [7]

Инверсия другого вида — инверсия слов — может быть, в принципе, обнаружена по тому, как человек использует язык. Например, иностранец, начиная изучать чужой для него язык, может путать некоторые слова. Так, человек, всю жизнь прошедший в Арктике, может не иметь устойчивых верований относительно эффекта "зелености" травы, "красности" пламени и т.д., он может так же, как и в предыдущем примере, называть зеленое красным и наоборот, но, в отличие от предыдущего случая, видеть красное зеленым и наоборот он не будет. Как тот же иностранец, путая слова "кувшин" и "чашка", на самом деле не верит в то, что люди за завтраком пьют из кувшина кофе, а не молоко. Инверсия слов носит временный характер, рано или поздно она проявляет себя в ходе диалога и устраняется при более совершенном овладении языком. Но из-за неоднозначности слов естественного языка и ситуационного характера правил их сочетаемости инверсия слов присутствует при любой степени

языковой компетенции (даже при очень хорошем владении языком она может возникнуть в области терминологической лексики). Можно вообразить другую ситуацию, подобную предыдущим, в которой гипнотизер заставляет испытуемого ощущать очень холодную воду как горячую. Тот, над кем производится этот эксперимент, будет по-прежнему называть горячей ту воду, которую таковой и чувствует, но факт инвертированности будет замечен окружающим. Если же на испытуемого воздействуют (например, посредством того же гипноза) так, что у него произойдет инверсия слов "горячее" и "холодное", то ситуация приблизится к той, что описана в первом примере, и происшедшую инверсию стимулов окружающим заметить будет сложно.

Инверсия стимулов не может быть устранена ни при какой степени языковой компетенции в силу так называемой гипотезы неискоряемости (*incorrigibility hypothesis*) [8, с.442], приписывающей особый статус нашим индивидуальным восприятиям и считающей отчеты о них никогда в принципе не могущими быть опровергнутыми как ошибочные кем-то другим. Допустим, например, что некто утверждает: "у меня болит зуб". Никто не может с полным основанием ответить на это "нет, ты ошибаешься". Факт здесь состоит в том, что говорящий не только уверен, что он испытывает боль (когда он ее испытывает), его уверенность не ждет подтверждения из каких-либо внешних по отношению к нему источников.

Если инверсия слов касается отчетов о восприятиях, то пока она не проявила себя как-либо в языковом поведении, она неотличима от инверсии стимулов и, следовательно, точно так же не обнаруживаема. Инверсия стимулов изначально заложена в самой природе участников диалога, инверсия же слов может возникнуть, например, при построении модели ограниченного естественного языка, допускающей хотя бы какую-то возможность нарушения своих ограничений. Если мы хотим построить модель, допускающую применение к ней требования осмысленности диалога, то в ней должны при-

существовать средства, позволяющие предъявителю текста воспроизвести любые возможные случаи инверсии, то есть либо описать их эксплицитно на имеющемся языке, либо изменить соответствующие им интерпретации. Отчеты о восприятиях эксплицитно описать на естественном языке нельзя: нельзя объяснить дальтонику, что такое красный цвет, так же, как человеку с нормальным зрением невозможно объяснить, что тот же дальтоник понимает под словом "красный". Интерпретацию в случае инверсии стимулов изменить тоже, скорее всего, нельзя; это можно сделать только в случае инверсии слов. Вопрос состоит в том, как в такой ситуации построить требуемую модель языка.

5. Инверсия может относиться либо к нашим восприятиям, либо к отображению нами этих восприятий в виде слов языка. Одним из общих положений психологии восприятия является утверждение о том, что восприятие есть не пассивное отражение наших впечатлений, это есть некоторое скрытое решение (принятое на внутреннем языке ума), оно основывается не только на текущих ощущениях, но и на всем предыдущем опыте. (Этим объясняется, в частности, возникновение иллюзий — еще один пример инверсии стимулов.) Мы имеем, следовательно, три ряда вещей: впечатления или ощущения (например, психическое состояние, соответствующее видению красного) → восприятия (внутренние суждения о типе ощущений) → слова естественного языка. Естественный язык устанавливает соответствие между словом и восприятием. Восприятие может быть представлено как некоторый предикат внутреннего (умственного) языка. Слово, выступающее в роли названия восприятия (например, слово "красный"), вообще не может возникнуть в отсутствии такого внутреннего предиката. Употреблению же этого слова предшествует установление истинности данного предиката, которая имеет место при наличии верования типа: $p =$ "верю, что я нахожусь в состоянии восприятия красного". Впечатление приравнивается, таким образом, к верованию, и образование та-

кого верования должно предшествовать возникновению соответствующего ему слова языка [7]. Это не означает, что данное мое впечатление тождественно тем, которые другие люди испытывают в тех же условиях (например, при виде красного цвета). Выражение p говорит лишь о том, каковым я внутренне решаю считать данное мое впечатление, используя для его наименования слово "красный".

В исчислении G слово "красный" может рассматриваться, например, как предикат вида $R(l, t)$ ("вижу красный цвет в месте l во время t "), интерпретация которого сопоставит символам R, l, t конкретные элементы предметной области M . Если следовать указанному взгляду на восприятие, то появлению предиката $R(l, t)$ в G предшествует образование во внутреннем языке U (который мы и можем считать уровнем "языковых умений" или метаязыком) предиката $R'(l, t)$, семантика которого должна допускать гипотезу несправимости. Этот предикат $R'(l, t)$ истинен, когда бы мы в него ни поверили. Язык U , на котором выражаются подобные верования, должен предшествовать обычному (внешнему) языку. Может появиться опасение относительно того, не возникнет ли и дальше такая ситуация, когда, проведя аналогичное рассуждение, придется постулировать еще один уровень, предшествующий уровню U ? Это не случится, так как в отношении уровня U мы можем сказать то, что нельзя было сказать для "внешнего" языка: что $R'(l, t)$ истинно, когда бы мы в него ни поверили. У нас нет необходимости предполагать наличие некоторого времени, предшествующего тому, когда мы вючили слово "R". Такая интерпретация предиката $R'(l, t)$ будет допускать и гипотезу несправимости.

6. Если мы будем строить модель, предусматривающую такое понимание предикатов типа $R(l, t)$, то при выходе на уровень U диалог можно будет трактовать не как обсуждение предметов, свойств, действий и т.д., а как обсуждение верований относительно предметов, свойств, действий и т.д. Разговор будет идти о состоянии веро-

вания, а не только о том, на что данное верование направлено. Это позволит предъявителю текста быть способным вообразить наличие у другого участника альтернативных верований. Первый участник диалога сможет принимать во внимание то, что объекты верований ему неизвестны и у него нет оснований судить о них с определенностью. Вводя же такую определенность, то есть выбирая объекты верований, он тем самым, в принципе, берет на себя ответственность за возможное предъявление не того текста, какой был нужен. В случае обнаружения инверсии слов верования могут быть изменены. Конечно, обращение к уровню *U* нужно только тогда, когда требования к точности моделирования диалога достаточно высоки.

Необходимость введения дополнительного метауровня *U* для построения диалогов, допускающих применение к ним требования осмысленности, оправдывает обращение к этому уровню в связи с моделированием естественного языка вообще. Метауровень включает средства, обеспечивающие такое специфическое свойство естественных языков как их самоизменяемость т.е. способность в ходе функционирования воздействовать на самих себя. Исследование других необходимых составляющих метауровня — предмет дальнейших публикаций.

Л т е р а т у р а

1. ТИМОФЕЕВА М.К. Языки действий //Методологические проблемы науки. — Новосибирск, 1991. — Вып.138: Вычислительные системы. — С.61-76.
2. ТИМОФЕЕВА М.К. Формальное описание языков действий //Теория вычислений и языки спецификаций. — Новосибирск, 1995. — Вып.152: Вычислительные системы. — С.38-60.
3. МОНТЕГЮ Р. Универсальная грамматика //Семантика и информатика. — 1985. — №26. — С.105-136.
4. SAINT-DIZIER P. An Approach to Natural Language Semantics in Logic Programming //The Journal of Logic Programming. — 1986. — Vol.3, №4. — P.329-356.

5. PUTNAM Hilary. Philosophy and Our Mental Life. //The Philosophy of Mind. — 1992. — P.91-99.
6. BLOCK Ned. Troubles with Functionalism //The Philosophy of Mind. — 1992. — P.69-90
7. LEEDS Stephen. Qualia, Awareness, Sellars //NOUS. — 1993.— Vol.27, № 3.— P.303-330.
8. ROBINSON Daniel N. An Intellectual History of Psychology. — The University of Wisconsin Press. — 1986.— 484 p.

Поступила в редакцию
19 декабря 1996 года