

Е. А. ГОЛОВИНА, К. Н. КОЛМЫЧЕК, В. Я. ТЕРЗИАН

ПРИНЦИПЫ ПРОВЕРКИ СЕМАНТИЧЕСКОЙ ПРАВИЛЬНОСТИ ЕСТЕСТВЕННОЯЗЫКОВЫХ ВЫСКАЗЫВАНИЙ

Анализ противоречивости естественных языковых высказываний. В естественных языковых системах в процессе анализа входные утверждения преобразуются к виду, позволяющему связать их со знаниями системы, или, другими словами, осуществляется понимание входных утверждений. В процессе такого преобразования может возникнуть противоречие между входными утверждениями и знаниями системы. Факт такого противоречия означает, что входное утверждение является неправильным (недопустимым) относительно знаний системы.

В диалоговой естественной языковой системе ДЕСТА [1] существуют два основных режима работы с системой: режим обучения и режим работы с пользователем. При работе в первом режиме система каждое входное утверждение считает заведомо правильным и заносит его в базу знаний. В режиме работы с пользователем система «с недоверием» относится к каждому входному утверждению и заносит его в базу знаний только в случае, если оно не противоречит ее знаниям.

Всякое входное утверждение представляется на внутреннем языке системы (*R*-представление [1]) в виде набора синтактико-семантических отношений (ССО) и семантических отношений

(СМНО) [1]. Отсюда вытекает существование двух типов противоречий: 1) противоречие в ССО; 2) противоречие в СМНО. Рассмотрим эти типы противоречий.

Противоречия в ССО возникают при несоответствии ССО входного утверждения по различным элементам ССО из базы знаний.

ССО в системе ДЕСТА в общем виде выглядит следующим образом: $B, A, C (HE, \bigcirc, D, D, E, E, X) = (HE, \bigcirc, D, D, E, E, X)$, где B — имя ССО; A — местоимение; C — предлог; HE — отрицание; \bigcirc — нормализованная словоформа; D — приставка; E — суффикс; X — предметная переменная.

Например факт: Петя не написал сочинение в классе — на языке R -представлений запишется в виде: $M1$: кто, он (не, писать, на) = (Петя); $M2$: что — делать (Петя) = (не, писать, на); $M3$: что — в (не, писать, на) = (сочинение); $M4$: где, в(не, писать, на) = (класс), где $M1, M2, M3, M4$ — метки ССО.

Возможны следующие виды противоречий в ССО:

1. Противоречия по предлогам. Они возникают в случае, если предлог в ССО данного предложения не совпадает с предлогом из соответствующего ССО базы знаний. Например, два ССО: где, на (лежать) = (стол); где, в(лежать) = (стол) являются противоречивыми.

2. Противоречия по отрицанию «не». Они возникают в случае, когда ССО базы знаний и ССО входного предложения различаются только наличием в одном из них частицы «не» и отсутствием этой частицы на соответствующем месте в другом ССО. Например, два ССО: кто, он (не, читать) = (Петя); кто, он(читать) = (Петя) являются противоречивыми.

3. Противоречия по словоформе. Здесь возможны два вида противоречий: а) противоречия по СМНО «противоположно» [1]. Они возникают, когда ССО базы знаний и ССО из входного утверждения отличаются парой словоформ, стоящих на соответственных местах, и словоформы этой пары связаны в базе знаний отношением «противоположно». Например, пара ССО: как (бежать) = (быстро); как (бежать) = (медленно) является противоречивой, так как противоположно (быстро) = (медленно); б) противоречия по классу. Они возникают, когда ССО базы знаний и ССО из входного утверждения отличаются парой словоформ, стоящих на соответственных местах, и словоформы этой пары являются элементами одного и того же класса. Например, пара ССО: какой (шар) = (голубой); какой (шар) = (красный) является противоречивой парой, а пара: какой (Петя) = (белокурый), какой (Петя) = (голубоглазый) не является противоречивой, так как элемент (цвет волос) = (белокурый), элемент (цвет глаз) = (голубоглазый).

4. Противоречия по приставке. Они возникают, если ССО базы знаний и ССО из входного утверждения отличаются приставками у соответствующих словоформ. В случае, если

приставки выражают одно и то же, противоречие не возникает. Например, пара ССО: что — делать (Петя) = (ехать, у); что — делать (Петя) = (ехать, при) является противоречивой, а пара: что — делать (Петя) = (ехать, до); что — делать (Петя) = (ехать, при) не является противоречивой.

5. Противоречия между словоформами данного ССО. Они возникают в том случае, если, согласно знаниям системы, объединение данной пары словоформ в некоторое ССО является недопустимым. Например, если системе известно: Снег может быть только белым, то ССО: какой (снег) = (красный) является противоречивым. Этот пример следует отличать от примера из случая 3(б). Если там входное ССО: какой (шар) = (красный) само по себе не является противоречивым, а входит в противоречие с ССО из базы знаний. Здесь же ССО является противоречивым само по себе.

Противоречия в СМНО возникают в случае несоответствия синтаксической структуры и смысла входных утверждений соответствующим СМНО базы знаний. Рассмотрим следующие виды таких противоречий.

Противоречия по СМНО «валентность» глагола [1]. Такие противоречия возникают в случае, когда морфолого-синтаксическая информация словоформ [1] входного утверждения не соответствует валентности глагола, выражающего действие во входном утверждении. Например, если системе известно, что валентность глагола «подарить»: {кто, кому, что — в}, то входное утверждение: «Петя подарил школе коллекцию марок» будет воспринято системой как противоречивое, поскольку словоформа «школе» отвечает на вопрос «чему» и не соответствует валентности.

Противоречия по СМНО «условие» [1]. Такие противоречия возникают в случае, когда для действия, о котором говорится во входном утверждении, не выполняются некоторые необходимые условия, заданные в СМНО «условие» данного действия.

Противоречия по СМНО «время». Они возникают в случае, когда абсолютное время реализации некоторой пары событий входит в противоречие с их относительной последовательностью. Например, следующий абстрактный текст: Ф1 раньше, чем Ф2. Ф1 в 1981 году. Ф2 в 1980 году, где Ф1, Ф2 — метки некоторых фактов, является противоречивым.

Все рассмотренные противоречия обнаруживаются в процессе вывода специальными процедурами, управляемыми соответствующими метапроцедурами.

Этапы семантического анализа естественных языковых выражений. Семантический анализ естественных языковых выражений (ЕЯВ) искусственной системой предназначен для их понимания, т. е. для установления связи некоторого входного ЕЯВ со знаниями системы. В процессе такого анализа устанавливается

возможное несоответствие, а иногда и противоречие между входными ЕЯВ и базой знаний [2].

В диалоговой естественной языковой системе ДЕСТА семантический анализ входных ЕЯВ проходит ряд последовательных этапов, соответствующих его различным уровням. При этом система переходит к более высокому уровню семантического анализа только в случае, если конкретная задача не может быть решена на более низком уровне. Существуют три уровня семантического анализа в системе ДЕСТА: 1) морфологический уровень семантического анализа; 2) синтаксический уровень семантического анализа; 3) семантический уровень семантического анализа. Рассмотрим на примерах работу подсистемы семантического анализа на этих трех уровнях.

1. На морфологическом уровне семантического анализа проверяется соответствие морфолого-синтаксической информации (МСИ) словоформ входного предложения валентности [1] глагола, выражающего действие из входного предложения. В режиме обучения каждое входное предложение принимается на веру. Поэтому валентность глаголов считается верной, и предложение сразу после второго просмотра и составления ССО записывается в С-структуру [1].

В режиме работы с пользователем проверка валентности начинается сразу же после второго морфологического просмотра [1], т. е. до составления ССО. Это позволяет значительно сократить время в случае обнаружения ошибки. В базе знаний валентность глагола уже сформировалась (может быть еще не полно) в режиме обучения. Необходимо учесть, что валентность глаголов с различными приставками разная. Например: Петя переплыл реку — валентность словоформы «переплыть» {кто, он; что — в}, Петя заплыл за дерево — валентность глагола «заплыть» — {кто, он; что, за}. Поэтому при проверке соответствия МСИ словоформ входного предложения валентности глагола необходимо приставку рассматривать как элемент, влияющий на валентность глагола.

С другой стороны, можно выделить три случая, возникающие при проверке валентности:

валентность глагола во входном предложении правильная, набор МСИ словоформ входного предложения в точности совпадает или является подмножеством валентности этого глагола с той же приставкой из базы знаний.

Валентность противоречива. Это может произойти, если а) соответствующие имена ССО из валентности и из МСИ словоформ входного предложения противоположны (типа: кто — что, кому — чему и т. п.); б) в соответствующих элементах валентности и МСИ имена ССО одинаковые, а местоимения разные или предлоги противоположны (типа: где, над и где, под; кто, он и кто, она). Но при этом надо учитывать, что могут

быть случаи, когда оба элемента таких пар входят в валентность. Тогда входное предложение правильно.

В остальных случаях мы только лишь можем сказать, что данный набор МСИ и валентность не соответствуют друг другу.

Рассмотрим алгоритм проверки валентности.

Ввод МСИ всех словоформ входного предложения.

Из этого набора МСИ отбрасываются те, которые соответствуют классам вопросов: что — делать и какой (так как словоформы с такими МСИ всегда не принадлежат валентности глагола).

После второго просмотра образуется массив, в котором каждой словоформе поставлены в соответствие абсолютный адрес словоформы в *S*-структуре, абсолютные адреса элементов МСИ и признак. По этому признаку отыскиваем глагол входного предложения.

Находим пересечение элементов строк, номерами которых являются абсолютный адрес словоформы и приставки. Тем самым мы выделяем те ССО, которые содержат данный глагол с приставкой.

Определяем факт пересечения полученного набора ССО с элементами строк, номерами которых являются абсолютные номера элементов МСИ для каждой словоформы в отдельности. Причем проверка проводится вначале по всем именам ССО, затем по местоимениям и лишь после этого по предлогам. Обработка полученных результатов проводится в соответствии с рассмотренными выше случаями. Здесь несоответствие возможно по всем трем компонентам МСИ: по вопросу, по местоимению, по предлогу. Рассмотрим примеры.

Пусть знания системы заданы следующим ЕЯВ: Петя подарил Маше книгу. МСИ для словоформ данного ЕЯВ будет соответственно: Петя {он, кто}; подарил {что — сделал}; Маше {она, кому}; книга {она, что — в}. Отсюда валентность глагола «дарить» — {он, кто; она, кому; она, что — в}. Пусть на вход системы поступит ЕЯВ: Петя подарил школе коллекцию. Сравнивая теперь МСИ для словоформ из данного ЕЯВ с валентностью глагола «дарить», обнаруживаем несоответствие по вопросу, так как МСИ «школе» — {она, чему}, а в валентности — {она, кому}.

Пусть знания системы те же, а на вход поступит ЕЯВ: Лена подарила Маше цветы. Здесь несоответствие по местоимению, так как МСИ «Лена» — {она, кто}, а в валентности — {он, кто; в} пусть знания системы заданы следующим ЕЯВ: Книга лежит на столе. МСИ для словоформ данного ЕЯВ будет соответственно: «книга» {она, что}; «лежит» {что — делает}; «столу» {он, где, на}. Отсюда валентность глагола «лежать» — {она, что; он, где, на}. Теперь входное утверждение: «Книга лежит в столе» не будет соответствовать валентности глагола «лежать» по предлогу.

2. Семантический анализ на синтаксическом уровне имеет место при составлении синтаксико-семантических отношений (ССО).

Пусть входное ЕЯВ будет: Петя читает книгу. МСИ для словоформ данного ЕЯВ соответственно: Петя {он, кто}; читает {что — делает}; книгу {она, что — в}. После нормализации [1] словоформ начинается формирование ССО. Сначала, используя МСИ, система делает заготовки для ССО: кто, он (X1) = (Петя); что — делать (X2) = (читать); что — в (X3) = (книга). Поиск соответствующих словоформ для X1, X2, X3 осуществляется по специальным правилам [1] и для данного примера приводит к однозначному конечному результату: M1: кто, он (читать) = (Петя); M2: что — делать (Петя) = (читать); что — в (читать) = (книга), где M1, M2, M3 — метки ССО. В случае многозначного решения вопроса о составлении ССО приходится переходить к следующему — семантическому уровню семантического анализа. Например, для ЕЯВ: «Заведующий лабораторией кивнул головой» мы не можем на синтаксическом уровне сказать, какую из словоформ: «заведующий» или «кивнуть» надо поставить на место X1 и X2 в заголовках для ССО данного примера: чем (X1) = (голова); чем (X2) = (лаборатория).

3. На семантическом уровне семантического анализа осуществляется проверка различных семантических отношений. Здесь проверяются такие СМНО, как условие, цель, назначение, зремя [1] и др.

Входное ЕЯВ будет семантически правильным, если между ним и фактами базы знаний в процессе анализа не возникли противоречия в приведенных СМНО [2].

Рассмотрим подробнее процесс анализа семантического отношения «условие». В режиме обучения в систему ДЕСТА заносятся необходимые и достаточные условия выполнения действий. Первые представляют собой минимальный набор условий, необходимых для того, чтобы некоторое действие могло иметь место. Например: Чтобы рубить, необходимо иметь чем рубить и иметь что рубить. Вторые представляют собой минимальный набор условий, достаточных для того, чтобы действие могло завершиться. Например: Чтобы срубить, достаточно рубить, иметь силу и время.

Для записи СМНО «условие» предусмотрена конструкция: условие (факт) = (список фактов).

Например, условием факта Ф1: «Петя подарил Маше цветы» является то, что Ф2: «У Пети есть цветы» и Ф3: «У Маши день рождения». Конструкция для условия будет выглядеть следующим образом: условие (Ф1) = (Ф1&Ф3).

В случае, если факт имеет сложную конструкцию или дополняется в ходе работы системы, список фактов, стоящий в правой части, пополняется и имеет более сложную конструкцию: условие (факт) = $(\bigvee_i (\bigwedge_j (\Phi_{ij})))$

Например, условием того, что Ф4: Петя имеет магнитофонные записи, является Ф5: Петя смотрел телевизор, Ф6: Петя записал на магнитофон песни, или Ф7: Петя взял магнитофонные записи у товарища, или Ф8: у Пети был день рождения, Ф9: Пете подарили хорошие песни. Тогда имеем: условие (Ф4) = $(\Phi5 \& \Phi6 \vee \Phi7 \vee \Phi8 \& \Phi9)$.

В естественных языковых высказываниях система ДЕСТА распознает семантическое отношение «условие» по конструкции «чтобы..., необходимо...».

Важным свойством для СМНО «условие» является транзитивность. Если для факта Ф1 условием является Ф2, а для Ф2 является Ф3, то Ф3 также является условием для Ф1. Из этого свойства в режиме умозаключения вытекает, что, если выполняется Ф1, то все его условия тоже выполняются, т. е. выполняются Ф2 и Ф3.

Для каждого входного высказывания процесс анализа СМНО «условие» состоит в следующем: система выясняет, выполняются ли на данный момент времени необходимые условия выполнения действия, о котором идет речь во входном предложении.

Остановимся непосредственно на алгоритме анализа семантического отношения «условие». После первого морфологического просмотра [1] анализируемого предложения каждой словоформе ставится в соответствие морфолого-синтаксическая информация, содержащая вопрос, на который отвечает данная словоформа. Это позволяет системе найти словоформу, обозначающую действие во входном предложении. Далее, в С-структуре система находит все множество (F1) словосочетаний, фактов, ситуаций, в которые входит данное действие и для которых существует семантическое отношение «условие», а также (F2) фактов, являющихся непосредственно условиями для элементов из F1. Затем система берет первый элемент из F1 и из слов входного предложения пытается составить, разрешая предметные переменные, синтактико-семантические отношения этого элемента. Так, путем перебора последовательно всех элементов из F1 находится тот из них, для которого данное соответствие устанавливается по всем ССО. Для данного элемента система находит соответствующие элементы из F2 при помощи семантического отношения «условие». Для них, используя значения предметных переменных, составляются ССО. Далее, факты, представленные этими ССО, в качестве общих вопросов поступают в подсистему синтеза ответа на вопрос. В случае утвердительного ответа на все поставленные вопросы условие действия считается выполненным и система переходит к формированию оставшихся ССО. Если же получен отрицательный ответ на один или несколько вопросов, то соответствующие условия считаются невыполненными и система сообщает об этом пользователю, указывая, какие именно условия не выполнены. И, на-

конец, при неопределенном ответе на некоторые вопросы система, используя элементы из $F2$, задает пользователю соответствующие вопросы о выполнении недостающих условий.

Рассмотрим работу алгоритма на конкретном примере. Пусть системе известно: Чтобы читать, необходимо уметь читать и иметь что читать. Петя умеет читать. Петя не имеет книгу. На языке ССО и СМНО это будет выглядеть следующим образом: $M1$: кто (читать) = $(X1)$; $M2$: что — делать $(X1)$ = (читать); $M3$: что — в (читать) = $(X2)$; $\Phi1 = M1 \& M2 \& M3$; $M4$: кто (уметь) = $(X1)$; $M5$: что — делать (уметь) = (читать); $M6$: что — делать $(X1)$ = (уметь); $\Phi2 = M4 \& M5 \& M6$; $M7$: кто (иметь) = $(X1)$; $M8$: что — делать $(X1)$ = (иметь); $M9$: что — в (иметь) = $(X2)$; $\Phi3 = M7 \& M8 \& M9$; условие $(\Phi1) = (\Phi2 \& \Phi3)$; $M10$: кто, он (уметь) = (Петя); $M11$: что—делать (Петя) = (уметь); $\Phi4 = M10 \& M5 \& M11$; $M12$: кто, он (иметь, не) = (Петя); $M13$: что — делать (Петя) = (иметь, не); $M14$: что — в (иметь, не) = (книга); $\Phi5 = M12 \& M13 \& M14$.

Пусть теперь на вход системы поступит следующее утверждение: Белокурый Петя читает интересную книгу товарищам в классе. После первого просмотра имеем: Белокурый (какой): Петя (кто, он); читает (что — делает); интересную (какую); книгу (что — в); товарищам (кому, они); класс (где, в). По вопросу «что делает» система находит действие «читает». По данному действию в S -структуре система находит множество $F1 = \{\Phi1\}$ и ему соответствующее множество $F2 = \{\Phi2, \Phi3\}$. Далее система пытается из слов входного предложения составить ССО факта $\Phi1$, разрешив предметные переменные соответственно: $X1 = \text{Петя}$; $X2 = \text{книга}$. Получаются следующие ССО входного предложения: кто, он (читать) = (Петя); что — делать (Петя) = (читать); что — в (читать) = (книга). Поскольку каждому ССО факта $\Phi1$ находится соответствующее ССО во входном предложении, то система из множества $F2$ берет факты $\Phi2$ и $\Phi3$, являющиеся условием факта $\Phi1$. Подставляя в ССО этих фактов значение разрешенных предметных переменных, получаем два набора ССО: {кто (уметь) = (Петя), что — делать (уметь) = (читать), что—делать (Петя) = (уметь)} и {кто (иметь) = (Петя), что — делать (Петя) = (иметь), что — в (иметь) = (книга)}, которые поступают в подсистему синтеза ответа на вопрос. Первый набор ССО целиком содержится в базе знаний (факт $\Phi4$), следовательно, на него дается утвердительный ответ. Второй набор ССО противоречит знаниям системы (факту $\Phi5$) по частице «не», следовательно, на него дается отрицательный ответ, т. е. условие «Петя имеет книгу» считается невыполненным. Система выдает пользователю факт базы знаний $\Phi5$ («Петя не имеет книгу»), вступивший в противоречие с входным предложением.

Далее необходимо остановиться на других семантических отношениях как на следствиях СМНО «условие».

Двойственным по отношению к СМНО «условие» является СМНО «результат». Оно указывает, какие могут быть результаты после выполнения некоторого действия [3].

Это семантическое отношение не является транзитивным, так как при выполнении данного факта мы не можем сказать, что выполняется и его результат, потому что для результата необходимо еще проверить и те факты, которые являются его условием. Таким образом, можно сказать, что СМНО «результат» только дает понятие о возможных последствиях действия. То есть, чтобы подвигаться в направлении «результат», необходимо подключить СМНО «условие» в режиме проверки семантической правильности и далее, если условия выполняются, можно принимать результат как завершенное действие.

В системе ДЕСТА СМНО «условие» и «результат» образуют структуру фактов, по которой в одном направлении можно двигаться с помощью СМНО «условие», в другом — с помощью СМНО «результат».

Непосредственно с семантическим отношением «результат» связано семантическое отношение «цель». Это СМНО так же осуществляет продвижение по структуре фактов подобно СМНО «результат», но «результат» связывает только непосредственно рядом стоящие в структуре факты, а «цель» связывает факты, стоящие друг от друга на несколько уровней. Это СМНО указывает, посредством каких «результатов» может быть достигнут тот или иной факт.

Эти основные семантические отношения для фактов, реализованные в системе ДЕСТА, являются важным инструментом проверки семантической правильности ЕЯВ и вывода умозаключений.

Список литературы: 1. *Ловицкий В. А.* Диалоговая естественная языковая система принятия решений. — Х.: Изд-во политехн. ин-та, 1981. — 110 с. 2. *Ловицкий В. А., Терзиян В. Я.* Семантический анализ в системе ДЕСТА. — В кн.: Интерактивные системы: Докл. и тез. докл. третьей школы-семинара. Тбилиси, 1980, с. 80—83. 3. *Терзиян В. Я.* Анализ, семантическая нормализация и идентификация естественных языковых текстов. — В кн.: Интерактивные системы: Тез. докл. и сообщ. четвертой школы-семинара. Тбилиси, 1982, с. 219—221.

Поступила в редколлегию 17.11.82.