

ПРИНЦИПЫ ОРГАНИЗАЦИИ АНАЛИЗА ЕСТЕСТВЕННОЯЗЫКОВЫХ ВЫСКАЗЫВАНИЙ В СИСТЕМАХ ОБЩЕНИЯ ПОЛЬЗОВАТЕЛЕЙ С ЭВМ.

СООБЩЕНИЕ 2

В сообщении [1] описаны основные теоретические положения, касающиеся оригинального подхода к организации анализа входной информации в системах естественного языкового общения, основанного на взаимодействии процедур различных видов анализа, управляемых метапроцедурами семантического анализа (СА). Процесс анализа было предложено разбить на ряд последовательных этапов в соответствии с валентностью входного естественного языкового высказывания (ЕЯВ), представляющей собой иерархическую структуру, построенную на основе множества валентностей словоформ [1] данного ЕЯВ. Этот процесс на каждом этапе рассматривается на трех уровнях собственно анализа: морфолого-синтаксическом (МСУ), синтактико-семантическом (ССУ), семантико-прагматическом (СПУ), которым предшествует нулевой уровень анализа (НУ), а также на трех уровнях понимания: информационно-синтаксическом (ИСУ), информационно-семантическом (ИСМУ), информационно-прагматическом (ИПУ) [1]. Цель данной организации анализа — попытка быстро решить конечную задачу анализа, т. е. выбрать необходимую реакцию системы на входное ЕЯВ, при возможном досрочном завершении процесса получения R -представления [2] для данного ЕЯВ.

Опишем работу метапроцедур СА на различных уровнях анализа входного ЕЯВ, использующих аппарат валентностей, процедуры вывода и формирования умозаключений, различного рода семантико-прагматические знания. Материал будет излагаться параллельно с разбором конкретных примеров.

Единицей R -представления в рассматриваемой системе является синтактико-семантическое отношение (ССО) [2]. Общий его вид

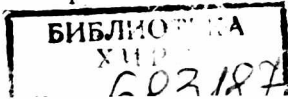
$$M_1: \text{МСИ}_2 (<\text{Словоформа 1}>) = (<\text{Словоформа 2}>),$$

где Словоформа 1 и Словоформа 2 — словоформы ЕЯВ, образующие синтагматическую пару; МСИ_2 — морфолого-синтаксическая информация второй словоформы; M_1 — метка ССО; $<>$ — признак нормы словоформы.

Например, ЕЯВ «Петя читает книгу» представляется в виде ССО следующим образом: M_1 : кто, он (читать) = (Петя); M_2 : что-делать (Петя) = (читать); M_3 : что-в, она (читать) = (книга).

Введем понятие обратной валентности словоформы.

Определение I. Обратной валентностью согласования



(ОВС) данной словоформы будем называть множество наборов МСИ, с которыми грамматически согласуется МСИ данной словоформы.

О п р е д е л е н и е 2. Обратной морфолого-синтаксической валентностью (ОМСВ) данной словоформы будем называть множество наборов МСИ, соответствующих словоформам, входящим слева от знака равенства в те ССО, в которые данная словоформа входит справа от знака равенства.

О п р е д е л е н и е 3. Обратной синтактико-семантической валентностью (ОССВ) данной словоформы будем называть множество словоформ, которые входят слева от знака равенства в те ССО, где данная словоформа расположена справа от знака равенства.

О п р е д е л е н и е 4. Обратной семантико-прагматической валентностью (ОСПВ) данной словоформы будем называть подмножество ОССВ, полученное из ССО, которые участвуют в фактах базы знаний, имеющих место в данный момент времени.

О п р е д е л е н и е 5. Обратной семантико-информационной валентностью (ОСИВ) данной словоформы будем называть подмножество ОССВ, полученное из ССО, относящихся к предметной области, которая является темой данного диалога.

О п р е д е л е н и е 6. Обратной прагматико-информационной валентностью данной словоформы будем называть множество словоформ, полученное в результате пересечения ОСПВ и ОСИВ. Обратную прагматико-информационную валентность (ОПИВ) будем также называть «обращенным» смыслом данной словоформы.

Наряду с описанными в предыдущем сообщении [1] валентностью согласования (ВС), морфолого-синтаксической валентностью (МСВ), синтактико-семантической валентностью (ССВ), семантико-прагматической валентностью (СПВ), семантико-информационной валентностью (СИВ) и прагматико-информационной валентностью (ПИВ) данной словоформы обратные валентности используются при анализе входного ЕЯВ.

Рассмотрим некоторые примеры.

1. ЕЯВ «Главный инженер проводит планерку» может быть представлена в виде ССО однозначно с использованием только ВС и ОВС: М1: кто, он (водить, про) = (инженер); М2: что-делать (инженер) = (водить, про); М3: что-в, она (водить, про) = (планерка); М4: какой (инженер) = (главный). Получив МСИ для словоформ с использованием ТВ-структуры [2], система сможет однозначно проанализировать на уровне валентностей согласования даже такой нонсенс: «Варкалось. Хливкие шорьки пырялись по наве. И хрюкотали зелюки как мюмзики в мове» (Л. Кэрролл «Алиса в зазеркалье»).

2. В ЕЯВ «Вано съел шашлык» на уровне ВС нельзя однозначно выделить подлежащее. Задача решается на МСУ, где будет использована МСВ для словоформы «съел»: {кто; что-в...}.

3. Однозначно определить подлежащее в ЕЯВ «Завод получил новый станок» можно только на ССУ, используя ССВ словоформы «получил».

4. Пусть системе известно: «Слесарь держит в руках напильник». Чтобы теперь при анализе следующего ЕЯВ: «Слесарь обрабатывает деталь» правильно сделать умозаключение, что «слесарь обрабатывает деталь напильником», необходимо привлечь семантико-прагматическую информацию: СПВ, использующую семантико-прагматическое отношение «условие» действия «обрабатывать деталь» [2].

При обработке анафорических ссылок в следующем фрагменте текста: «Волк напал на охотника. Он его застрелил», задача однозначно решается также с помощью СПВ, использующей условие действия «застрелить».

5. В случае анализа ЕЯВ «К завтрашнему дню жених попросил наряд», чтобы сделать умозаключение, где был спрошен наряд: у бригадира на работу, или в ателье мод, необходимо сначала определить предметную область, к которой относится ЕЯВ, используя словоформу «жених», а затем применить СИВ, соответствующую данной предметной области.

6. При анализе фрагмента текста «Писатель просматривал за столом текст рукописи. Он его отшлифовал» для разрешения анафорической ссылки: «его — стол» или «его — текст» требуются и знания о конкретной ситуации, и знания о предметной области. Поэтому в данном случае будет использована ПИВ.

Рассмотрим примеры простейших умозаключений [3].

1. Пусть системе известно: «Петя читает книгу». Пусть теперь на вход системы поступает ЕЯВ «Петя читает быстро». Полученный в результате анализа данного ЕЯВ факт не имеет смысла заносить целиком в память системы. Вместо этого процедура формирования умозаключений находит в базе знаний факт: «Петя читает книгу» и дописывает к нему ССО из анализируемого факта: как (читать) = (быстро). В результате работы процедуры в базе знаний системы будет храниться факт: «Петя читает книгу быстро».

2. Пусть системе известно: «Петя имеет книгу». На вход поступает ЕЯВ «Петя читает». Чтобы ответить на вопрос: «Что читает Петя?» процедура формирования умозаключений использует СПВ словоформы «читать», учитывающую проверку возможности выполнения данного действия в текущей ситуации.

Рассмотрим некоторые правила извлечения из базы знаний и работы с МСВ и ССВ. При извлечении из базы знаний системы МСВ необходимо учитывать взаимосвязь элементов валентности, поскольку возможны случаи, когда один элемент с необходимостью влечет другой. Другими словами, валентность — это не просто прямое объединение МСИ, а объединение с учетом связей МСИ в фактах базы знаний. Например, если в базе знаний содержатся факты: «Маша любит Васю; Петя любит Лену»,

то валентность «любить» не равна множеству: {кто, она; кого он; кто, он; кого, она}, а равна объединению в два различных множества пар МСИ, принадлежащих разным фактам: {кто, она кого, он} и {кто, он; кого, она}. Для выяснения того, в каких случаях при извлечении МСВ из базы знаний необходимо разбивать множество МСИ по фактам, предназначено

Правило 1. Если в наборе валентных МСИ для данной словоформы имеются противоположные пары вопросительных слов (кто — что, кому — чему и т. д.) или одному и тому же вопросительному слову соответствуют различные местоимения (кто, он — кто, она; кому, они — кому, оно и т. п.), то необходимо сначала произвести разбиение данного набора МСИ по следовательно по каждой из этих противоположностей, объединяя МСИ соответствующих фактов и получая два или более множества МСИ. Теперь, отбросив из этих множеств те элементы, по которым осуществлялось разбиение, находится симметрическая разность данных множеств. Если эта разность содержит противоположные МСИ, принадлежащие различным исходным множествам, то эти множества являются разбиением валентности данной словоформы по данной противоположности. Если эта разность ни для одной противоположности не будет содержать указанных МСИ, то валентность данного глагола можно брать как простое объединение валентных МСИ.

Правило 2. Если в наборе валентных МСИ для некоторого глагола имеются МСИ класса: «что-дела...», то МСВ представляется в виде двух множеств M_1 и M_2 , где M_1 — набор валентных МСИ из всех фактов, содержащих МСИ указанного класса и участвующих в формировании валентности данного глагола, а M_2 — набор оставшихся валентных МСИ. К M_2 должно быть применено правило 1.

На ССУ анализа разрешаются неоднозначности, возникшие на МСУ, путем проверки допустимости на ССУ пар, полученных на МСУ. Анализ на ССУ производится с использованием ССВ по определенным правилам в несколько этапов. Переход к каждому следующему этапу происходит в случае неудачи на предыдущем. Каждое правило имеет свой коэффициент, который затем присваивается паре, если она удовлетворяет изложенному в правиле требованию. Чем выше коэффициент, тем точнее можно сказать, что данная пара семантически верна на ССУ.

Правило 3 (коэффициент $K=10$). Если в базе знаний есть ССО, в точности соответствующее данной паре, то последняя считается допустимой на ССУ.

Пусть общий вид некоторого ССО базы знаний в несколько упрощенной форме будет выглядеть: $M(X_1) = (X_2)$, где M — имя ССО, X_1 и X_2 — словоформы.

Правило 4. Если для пары словоформ Y_1, Y_2 имеет место $X_2 = Y_2$
и а) Y_1 — синоним X_1 ($K=9$); или

- б) У1 — элемент класса X1 ($K=6$); или
- в) X1, У1 — элементы одного класса ($K=3$),

то пара считается допустимой на ССУ.

П р а в и л о 5. Если для пары словоформ У1, У2 имеет место:

$$X1 = Y1$$

- и а) У2 — синоним Х2 ($K=8$); или
- б) У2 — элемент класса Х2 ($K=5$); или
- в) Х2, У2 — элементы одного класса ($K=2$),

причем Х2 — не прилагательное, то пара считается допустимой на ССУ.

П р а в и л о 6. Если для пары словоформ У1, У2 имеет место

- а) У1 — синоним Х1 ($K/2=3,5$); или
- б) У1 — элемент класса Х1 ($K/2=2$); или
- в) Х1, У1 — элементы одного класса ($K/2=0,5$)
- и г) У2 — синоним Х2 ($K/2=3,5$); или
- д) У2 — элемент класса Х2 ($K/2=2$); или
- е) Х2, У2 — элементы одного класса ($K/2=0,5$),

причем Х1 — не глагол, а Х2 — не прилагательное, то пара считается допустимой на ССУ.

Последние три правила для случаев, когда а) Х1 — элемент класса У1; б) Х2 — элемент класса У2 имеют коэффициент 0,5.

В случае, если ни одно из правил не подходит, $K=0$.

Как уже говорилось, на основе прямых и обратных валентностей осуществляется объединение словоформ входного ЕЯВ в ССО. Правило объединения словоформ в ССО с использованием ВС или МСВ имеет следующий вид.

П р а в и л о 7. Если для двух словоформ, принадлежащих входному ЕЯВ, выполняются следующие условия: а) некоторый вариант МСИ второй словоформы принадлежит МСВ (ВС) первой словоформы; б) МСИ первой словоформы принадлежит ОМСВ (ОВС) второй словоформы, причем обратные валентности берутся по тому же набору МСИ, по которому были взяты прямые; в) ни у одной из оставшихся словоформ входного ЕЯВ не существует такого варианта МСИ, который бы принадлежал ОМСВ (ОВС) второй словоформы, то данная пара словоформ может быть в соответствующем порядке объединена в ССО, причем именем этого ССО будет являться тот вариант МСИ второй словоформы, который удовлетворяет условию а).

В случае выполнения только условий а и б данного правила сохраняется многозначность, которая может быть разрешена на более высоком уровне анализа.

Правило объединения словоформ в ССО с использованием ССВ, СПВ, СИВ, ПИВ имеет следующий вид.

П р а в и л о 8. Если для пары словоформ входного ЕЯВ, которая является допустимой на предыдущем уровне анализа по условиям а и б, выполняются следующие условия: а) нормализованная словоформа, соответствующая второй словоформе данной пары, принадлежит валентности первой словоформы по

допустимому на МСУ варианту МСИ этой словоформы; б) нормализованная словоформа, соответствующая первой словоформе данной пары, принадлежит обратной валентности второй словоформы по допустимому на МСУ варианту МСИ этой словоформы; в) ни одна из оставшихся словоформ входного ЕЯВ, согласующихся со второй словоформой рассматриваемой пары на предыдущем уровне анализа, не имеет нормы, принадлежащей обратной валентности второй словоформы по допустимому на МСУ варианту МСИ этой словоформы, то данная пара словоформ может быть в соответствующем порядке объединена в ССО, причем именем этого ССО будет являться тот вариант МСИ второй словоформы, по которому была выявлена допустимость рассматриваемой пары на предыдущем уровне анализа.

Для того чтобы полученное в результате анализа входного ЕЯВ внутреннее представление было удобно для вопросно-ответного режима и позволяло семантически однозначно синтезировать выходные ЕЯВ, применяется так называемая операция свертки. Правило использования операции свертки имеет следующий вид.

П р а в и л о 9. Если в результате анализа входного ЕЯВ некоторая словоформа, не являющаяся сказуемым ЕЯВ, входит слева в два и более ССО, то эти ССО объединяются меткой ситуации, после чего данная метка записывается в правую часть тех ССО, где исходная словоформа располагалась справа. При этом образуются новые ССО входного ЕЯВ, а операция их получения называется операцией свертки.

Например, результатом анализа ЕЯВ «Веселый голубоглазый мальчик в белой сатиновой рубашке быстро бежал по мягкой росистой траве» будет следующий набор ССО:

- М1: КТО, ОН (БЕЖАТЬ) = (МАЛЬЧИК)
- М2: ЧТО-ДЕЛАТЬ (МАЛЬЧИК) = (БЕЖАТЬ)
- М3: КАК (БЕЖАТЬ) = (БЫСТРО)
- М4: ГДЕ, ОНА, ПО (БЕЖАТЬ) = (ТРАВА)
- М5: КАКОЙ (МАЛЬЧИК) = (ВЕСЕЛЫЙ)
- М6: КАКОЙ (МАЛЬЧИК) = (ГОЛУБОГЛАЗЫЙ)
- М7: ЧЕМ, ОНА, В (МАЛЬЧИК) = (РУБАШКА)
- М8: КАКОЙ (РУБАШКА) = (БЕЛЫЙ)
- М9: КАКОЙ (РУБАШКА) = (САТИНОВЫЙ)
- М10: КАКОЙ (ТРАВА) = (МЯГКИЙ)
- М11: КАКОЙ (ТРАВА) = (РОСИСТЫЙ)

В результате применения к приведенному набору ССО операции свертки получаем следующий результат:

- МН1: КТО, ОН = (МАЛЬЧИК)
- МН2: ГДЕ, ОНА, ПО = (ТРАВА)
- МН3: ЧЕМ, ОНА, В = (РУБАШКА)
- СЛСЧ1: МН1 и М5 и М6
- СЛСЧ2: МН2 и М10 и М11
- СЛСЧ3: МН3 и М8 и М9
- СЛСЧ4: СЛСЧ1 и М12

М12: ЧЕМ, ОНА, В (МАЛЬЧИК) = (СЛСЧ3)

М13: КТО, ОН, (БЕЖАТЬ) = (СЛСЧ4)

М14: ГДЕ, ОНА, ПО (БЕЖАТЬ) = (СЛСЧ2)

Ф1: М13 и М2 и М3 и М14,

где М — метка ССО; МН — метка неполного ССО; СЛСЧ — метка словосочетания; Ф — метка факта.

Рассмотрим процесс анализа входных ЕЯВ на конкретном примере. Пусть на вход системы, в которой уже имеется некоторый набор знаний, поступило ЕЯВ «Кивнул головой заведующий лабораторией заведующей столовой». Анализ начинается с того, что каждой словоформе ЕЯВ ставятся в соответствие допустимые варианты МСИ, включающей в себя вопрос, на который отвечает данная словоформа, местоимение, соответствующее норме словоформы, предлог, относящийся к данной словоформе. Для приведенного примера имеем: МСИ1 (кивнул) = {что-сделал}; МСИ2 (головой) = {чем, она}; МСИ3 (заведующий) = {кто, он}; МСИ4 (лабораторией) = {чем, она}; МСИ5 (заведующей) = {кого, она; кому, она; кем, она}; МСИ6 (столовой) = {чего, она; чему, она; чем, она}. Теперь система приступает к первому этапу анализа ЕЯВ. Главные члены ЕЯВ выбирают однозначно уже на МСУ и для них формируются ССО.

М1: КТО, ОН (КИВНУТЬ) = (ЗАВЕДУЮЩИЙ)

М2: ЧТО-СДЕЛАЛ (ЗАВЕДУЮЩИЙ) = (КИВНУТЬ)

На МСУ второго этапа из входного ЕЯВ выбираются словоформы, кроме подлежащего, которые согласуются со сказуемым на уровне МСИ. Возможные согласования: а) МСИ1 — МСИ2; б) МСИ1 — МСИ4; в) МСИ1 — МСИ5; г) МСИ1 — МСИ6. Возникла неоднозначность по вопросу «чем(кем)». Осуществляется переход к ССУ анализа. С помощью валентности глагола «кивнуть» минимизируется МСИ: из МСИ5 отбрасывается набор «кого, она» и «кем, она», а из МСИ6 — «чего, она». На ССУ анализа проверяется допустимость пар: «кивнул — головой», «кивнул — лабораторией», «кивнул — заведующей» и «кивнул — столовой». Допустимыми являются только первая и третья пары, для которых составляются ССО

М3: ЧЕМ, ОНА (КИВНУТЬ) = (ГОЛОВА)

М4: КОМУ, ОНА (КИВНУТЬ) = (ЗАВЕДУЮЩИЙ)

На третьем этапе анализа находятся возможные согласования подлежащего с еще не задействованными в ССО словоформами ЕЯВ: а) МСИ3 — МСИ4; б) МСИ3 — МСИ6. Снова неоднозначность, которая разрешается по правилу «соседства» [2]. Имеем ССО

М5: ЧЕМ, ОНА (ЗАВЕДУЮЩИЙ) = (ЛАБОРАТОРИЯ)

На последнем этапе находится согласование для дополнения «заведующей», которое единственно: МСИ5 — МСИ6. Имеем ССО

М6: ЧЕМ, ОНА (ЗАВЕДУЮЩИЙ) = (СТОЛОВАЯ)

Легко видеть, что из приведенного набора ССО помимо исходного ЕЯВ можно синтезировать ЕЯВ «Заведующий столовой кивнул заведующей лабораторией». Чтобы не допускать подобных неоднозначностей, необходимо произвести операцию свертки по следующему правилу.

П р а в и л о 10. Если две различные словоформы некоторого ЕЯВ имеют одну и ту же норму, которая в результате анализа данного ЕЯВ оказывается в левой части двух различных ССО, имеющих одно и то же имя, то относительно этих словоформ необходимо произвести операцию свертки, которая осуществляется аналогично предыдущему правилу.

Для рассматриваемого примера в результате операции свертки имеем:

МН1: КТО. ОН = (ЗАВЕДУЮЩИЙ)
МН2: КТО. ОНА = (ЗАВЕДУЮЩИЙ)
СЛСЧ1: МН1 и М5
СЛСЧ2: МН2 и М6
М7: КТО. ОН (КИВНУТЬ) = (СЛСЧ1)
М8: ЧТО-СДЕЛАЛ (СЛСЧ1) = (КИВНУТЬ)
М9: КОМУ. ОНА (КИВНУТЬ) = (СЛСЧ2)
Ф1: М7 и М8 и М9

Таким образом, приведенная организация анализа предполагает немедленное разрешение возникающих неоднозначностей и кроме этого подготавливает всю необходимую информацию для осуществления семантически однозначного синтеза выходных ЕЯВ.

Список литературы: 1. Терзиян В. Я. Принципы организации анализа естественных языковых высказываний в системах общения пользователей с ЭВМ. *Сообщение I*. — Пробл. бионики, 1984, вып. 34, с. 47—51. 2. Ловицкий В. А. Диалоговая естественная языковая система принятия решений — Х.: Вища шк. Изд-во при Харьк. ун-те, 1981. — 110 с. 3. Терзиян В. Я. Вывод умозаключений из естественных языковых высказываний. — В кн.: Проблемно-ориентированные диалоговые комплексы: Тез. докл. Кишинев, 1983, с. 102—104.

Поступила в редколлегию 02.04.84.

УДК 510. 62

Т. Г. КАЛЕКИНА

О МАТЕМАТИЧЕСКОМ ОПИСАНИИ ЛИНГВИСТИЧЕСКИХ ЗАКОНОВ ЗВУКОВЫХ СОЧЕТАНИЙ

Современный подход к автоматическому пониманию речи заключается в привлечении лингвистической информации для полного распознавания речевого сообщения. Аналогично тому, как человек использует в естественном речевом общении свое знание языка, системам автоматического понимания речи необхо-