УДК 62.506.2

И. В. КОМАРКОВА, В. Я. ТЕРЗИЯН

АЛГОРИТМ РАЗРЕШЕНИЯ МЕСТОИМЕНИЙ В СИСТЕМАХ МАШИННОГО АНАЛИЗА ТЕКСТОВ

Важной задачей семантического анализа текстов является задача разрешения местоимений, т. е. замены их соответствующими объектами [1].

Естественноязыковая информация поступает в естественноязыковую систему по-разному: отдельными фактами, а чаще — совокупностью фактов (текст). Если в первом случае объекты, упомянутые в предложении, должны быть строго определены: ЛЕС, КОЛЯ, СТОЛ, КНИГА и т. д., то во втором случае те или иные объекты при неоднократном упоминании о них могут заменяться словозаменителями — местоимениями. Но в каждом случае появляется вопрос: какие же объекты заменяют каждое местоимение? Прежде чем ответить на этот вопрос, дадим следующую классификацию местоимений [2]: 1) личные: я, ты, он, она, они, мы, вы, ..., 2) притяжательные: мои, ваш, его, твой, их ...; 3) указательные: этот, тот, туда, оттуда, ...; 4) возвратные: себя, свой, ...; 5) порядковые: один, другой, первый, второй, ... (в случае, когда они не являются числительными).

Пока мы считаем достаточным ограничиться рассмотрением этих пяти классов. Коротко охарактеризуем местоимения каждого класса.

1. *Личные.* Характеристики: число, род, одушевленность соответствующего объекта (эти характеристики, если они могут быть определены, ускоряют процесс поиска эквивалента данного местоимения, причем одушевленность, как правило, определяется по семантике глагола).

Тип эквивалента: имя существительное.

Эквивалентом местоимения *я (мы)* является (являются) лицо (лица), ведущее (ведущие) рассказ. Например: «И я проснулся», — закончил Петя».

Аналогично эквивалентом местоимения *ты (вы)* является (являются) лицо (лица), к которым обращается рассказчик. Например: «Люблю тебя», — шепнул Матвей Вале».

Эквивалентами же местоимений *он, она, оно, они* могут служить любые объекты или объекты, упомянутые в тексте, и соответствие между местоимением и его эквивалентом устанавливается путем семантического анализа текста посредством определенного метода. Например: «Петя вошел в дом. Он разделся и сразу же лег спать».

Естественно, никакой анализ не сможет найти эквивалент для местоимения в случае, когда это бессилен сделать даже человеческий мозг. С этим явлением мы сталкиваемся в подобных случаях: «Маша подарила Оле цветы. Ей было очень приятно». Трудно сказать: кому же все-таки было приятно?

Следует также отметить тот факт, что одно и то же местоимение может в одном и том же тексте (реже в одном и том же предложении) иметь различные эквиваленты. Например: «Коля взял у Кати книгу. Он прочитал ее книгу. Коля отдал ее Пете». Одно и то же местоимение «ее» в данном тексте имеет два разных эквивалента: «Катя» и «Книга».

2. *Притяжательные.* Характеристики: число, род, одушевленность.

Тип эквивалента: имя существительное.

Данный класс является производным от класса личных местоимений. Для этого класса местоимений заметим, что местоимения *его, ее, их* в одних случаях являются личными, в других — притяжательными. Например: «Коля хорошо пел. Его любили слушать». Его — личное местоимение. «Коля встретил Катю. Она отдала его книгу Пете». Его — притяжательное местоимение.

3. *Указательные.* Характеристики: число, род.

Этот класс отличается от других тем, что конкретизирует, т. е. выделяет из общей совокупности объектов, имеющих одно имя, именно тот, который либо сам совершал действие, либо над ним совершали действие. Например: «Лена играла вальс Шопена. Эта девочка бесспорно талантлива». Или: «Петя достал с полки книгу. Он отдал эту книгу Коле».

4. *Возвратные.* Характеристики: число, род. Эквивалентом является во всех случаях субъект, выполняющий действие. Например: «Коля взял для себя портфель в шкафу».

5. *Порядковые.* Характеристики: число, род, одушевленность.

Тип эквивалента: существительное, прилагательное. Причем общий эквивалент обязательно состоит из нескольких объектов единственного числа или является множественным числом.

Например: «Петя и Вася закончили школу. Один с золотой медалью, другой — без». Общим эквивалентом является «Петя+Вася». Возможны две комбинации: «Петя кончил с медалью, Вася — без» и «Вася кончил с медалью, Петя — без». Оба варианта семантически верны. Если бы наш пример изменить: «Петя и Вася закончили школу. Первый с медалью, Вася— без», то эквивалент «первый» — «Петя».

В случае, когда эквивалент представлен множественным числом, мы в процессе разрешения данные местоимения переводим в числительные.

Например: «Дети пошли в школу. Один раньше, другой позже». Общий эквивалент дети, множественное число, одушевленный. Но в нашем тексте не уточняется «состав» слова «дети»

и в этом случае мы, пользуясь ТВ-структурой (структурой морфологического анализа) [3], преобразуем множественное число «дети» в единственное «ребенок» и переводим местоимения в числительные: «Дети пошли в школу. Один ребенок раньше, другой ребенок позже».

Может быть случай, когда сами местоимения 5-го класса встречаются во множественном числе. Например: «Дети шли на демонстрацию. Одни с флажками, другие с шариками». В этом случае мы общий эквивалент «дети» оставляем без изменения, а местоимения преобразуем в числительные: «Дети шли на демонстрацию. Одни дети с флажками, другие дети с шариками».

Рассмотрим общий алгоритм разрешения местоимений, т. е. однозначного определения их эквивалентов, разработанный на базе действующей естественноязыковой системы ДЕСТА [3]. 1. Выделение всех местоимений в предложении. 2. Определение типа местоимений и их характеристик. 3. Определение типа эквивалента. 4. Проверка на одновременную уместность (выживаемость). 5. Проверка на порядок. 6. Просмотр текста от последнего местоимения к началу и выбор нужного числа эквивалентов. 7. Определение возможных подстановок эквивалентов. 8. Подключение подсистемы понимания и проверки на истинность каждой возможной подстановки эквивалентов. Если подсистема понимания выявит противоречие в предложении с данным вариантом эквивалента, то переходим к п. 9. В противном случае система переходит к следующему местоимению (если оно есть); если его нет, то продолжает работу над текстом. 9. Переходим к другому варианту эквивалента. Перейти к п. 8.

Сделаем некоторые пояснения по данному алгоритму.

1) Выделение местоимений производится при помощи поиска в словаре нетерминальных символов [3], в котором местоимения выделены в отдельный класс.

2) Определение типа местоимений определяется также по словарю нетерминальных символов, в котором местоимения разбиты на подклассы.

Характеристики местоимений (или морфолого-синтаксическая информация) определяются при помощи ТВ-структуры, через которую вместе с другими словоформами пропускаются и местоимения. Одушевленность или неодушевленность определяется на уровне семантического анализа по валентности (т. е. набору вопросов, на которые отвечают участники действия) глагола. Например: «Петя рубит тополь. Он его рубит уже полчаса». Валентность (рубить) = (кто, что-в). Следовательно, местоимение «он» относится к одушевленному объекту, а «его» — к неодушевленному, т. е.: «Петя тополь рубит уже полчаса».

3) Определение типа эквивалентов и их характеристик производится с целью замены местоимений допустимыми по типу и характеристикам эквивалентами. Тип и характеристики эквивалентов определяются с помощью ТВ-структуры.

4) Проверка на одновременную уместность означает, что если в предложении встречаются личные местоимения с притяжательными или возвратными, то необходимо проверить, не соответствует ли парам: личное, возвратное; личное, притяжательное один эквивалент. Например: «Коля с Машей встретились в классе. Он отдал ей ее портфель и забрал свою ручку».

5) Проверка на порядок. Если в сложноподчиненном предложении есть конструкции типа:

ТОТ ..., КОТОРЫЙ ...
(согласуются в числе и роде)
ТУДА ..., ГДЕ ...
ТУДА ..., КУДА ...
ОТТУДА ..., ОТКУДА ...,

то местоимениям «тот», «туда», «оттуда» ставятся в соответствие те эквиваленты, которые обладают признаками, перечисленными после слов «который», «где», «куда», «откуда». Если в предложении встретится возвратное местоимение, то ему соответствует объект, выполняющий действие. Например: «Петя отдал Коле свою ручку» или «Коле свою ручку дал Петя» (т. е. независимо от того, где стоит местоимение — до или после эквивалента).

6) Выбор нужного числа эквивалентов определяется по числу местоимения (единственное или множественное): т. е. если местоимение у нас единственного числа, то, естественно, ему будет соответствовать один эквивалент также в единственном числе.

Если же местоимение во множественном числе, то эквивалентом его может служить:

а) один объект или субъект также во множественном числе;

б) несколько объектов или субъектов в единственном или множественном числе.

Например: «Коля учится в 10-А классе. Лена учится в 10-А классе». Они сидят за одной партой». Эквивалентом местоимения ОНИ служит «Коля+Лена».

Другой пример: «Котята не умели плавать. Все они утонули». Эквивалент местоимения ОНИ — «котята». Или «Дубы, березы, клены ... Они радовали глаз горожанина».

7) Определение возможных подстановок эквивалентов говорит о том, что характеристики эквивалентов в каждой подстановке должны соответствовать характеристикам местоимений.

8) Каждая возможная подстановка проверяется на уровне семантического анализа и считается допустимой, если понимается системой на синтаксическом и семантическом уровнях понимания.

Возможны случаи, когда на выходе после проверки на «смысл» получаются две или более истинных гипотезы. Например: «Встретились Петя и Вася. Он поздравил его с днем рож-

дения». Обе подстановки: «он — Петя, его — Васю и он — Вася, его — Петю» семантически верны.

Рассмотрим работу системы на примере анализа личных местоимений 3-го лица единственного и множественного числа: «он», «она», «оно», «они» в различных падежах.

Все местоимения в системе ДЕСТА относятся к нетерминальным символам и образуют в NTV-словаре, т. е. словаре нетерминальных символов, отдельный класс.

Рассмотрим алгоритм работы процедуры разрешения личных местоимений на конкретном примере. Пусть в систему введен текст: «Оля имеет вазу. Лена подарила Оле цветы. Она поставила их в вазу». Прежде, т. е. до момента получения этого примера, в базе знаний системы хранились в виде R-представлений [3] такие знания:

Чтобы подарить, необходимо иметь, что дарить. М1 : кто (дарить, по) = (X1); М2: что-д. (XI) = (дарить, по); М3: что-в (дарить, по) = (X2). Ф1 = М1&М2&М3. М4: кто (иметь) = = (X1); М5: что-д. (X1) = (иметь); М6: что-в (иметь) = (X2); Ф2=М4&М5&М6. усл. (Ф1) = (Ф2), где M_i—метки синтактико-семантических отношений (ССО), Φ_i — метки фактов, усл. — семантическое отношение (СМНО) «условие», X_i — предметные переменные [3].

В результате действия подарить тот, кто дарит, не имеет то, что дарит, а тот, кому дарит, имеет то, что дарит. М6: кому (дарить, по) = (X3); Ф3 = М1&М2&М3&М6; М7: кто (иметь, не) = (X1) М8: что-д. (X1) = (иметь, не); М9: что-в (иметь, не) = (X2). Ф4 = М7&М8&М9; М11: кто (иметь) = (X3); М12: что-д. (X3) = (иметь). Ф5 = М11&М12М5. Рез. (Ф3) = (Ф4&Ф5), где рез. — СМНО «результат».

Чтобы что-то куда-то поставить, необходимо иметь то, что поставить, и иметь куда поставить: М13: кто (ставить, по) = (X4); М14: что-д. (X4) = (ставить, по) М15: что-в (ставить, по) = (X5); М16: куда, в (ставить, по) = (X6). Ф6: М13&М14&М15&М16. М17: кто (иметь) = (X4) М18: что-д. (X4) = (иметь); М19: что-в (иметь) = (X5); М20: что-в (иметь) = (X6). Ф7=М17&М18&М19. Ф8=М17&М18&М20. Усл. (Ф6) = (Ф7&Ф8).

В результате действия поставить что-то во что-то то, что ставят, находится в том, куда ставят. М21: что (находиться) = (X5); М22: что-д. (X5) = (находиться); М23: где, в (находиться) = (X6). Ф9=М21&М22&М23. Рез. (Ф6) = (Ф9).

Когда в систему поступил пример, она на основе имеющейся базы знаний приступает к анализу текста и преобразует в R-представление: «Оля имеет вазу». М30: кто, она (иметь) = (Оля); М31: что-д. (Оля) = (иметь); М32: что-в (иметь) = (ваза). Ф10=М30&М31&М32.

Затем система анализирует второе предложение: «Лена подарила Оле цветы». М33: кто, она (дарить, по) = (Лена). М34:

что-д. (Лена) = (дарить, по); М35: что-в (дарить, по) = (цветы); М36: кому (дарить, по) = (Оля). Ф11 = М33&М34&М35&М36.

Сравнивая с базой знаний Ф11, система делает вывод, что: «Лена не имеет цветы». «Оля имеет цветы», т. е. М37: кто, она (иметь, не) = (Лена). М38: что-д. (Лена) = (иметь, не). М39: что-в (иметь, не) = (цветы). Ф12 = = М37&М38&М39&. М30: кто, она (иметь) = (Оля). М31: что-д. (Оля) = (иметь). М42: что-в (иметь) = (цветы). Ф13 = М30& М31&М42, т. е. рез. (Ф11) = = (Ф12&Ф13).

И, наконец, система переходит к анализу третьего предложения.

Используя алгоритм разрешения местоимений, система делает вывод, что эквивалент местоимения «их» есть «цветы». Эквивалентом же местоимения «Она» могут служить два существительных: «Оля» и «Лена». Но, используя семантическое отношение условие, определенное в базе знаний для действия «поставить», система приходит к выводу, что этому отношению удовлетворяет вариант: «Оля поставила цветы в вазу».

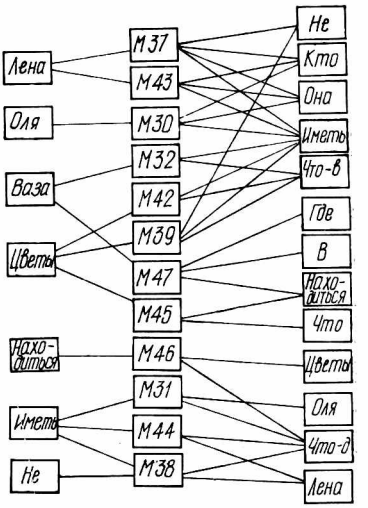

Рис. 1

Весь текст во внутреннем представлении будет выглядеть в виде набора ситуаций, имеющих место в последовательные моменты времени: Т0: Лена имеет цветы. Оля имеет вазу.

Т1: Лена не имеет цветы. Оля имеет цветы.

Т2: Лена не имеет цветы. Оля имеет цветы. Оля имеет вазу. Цветы находятся в вазе.

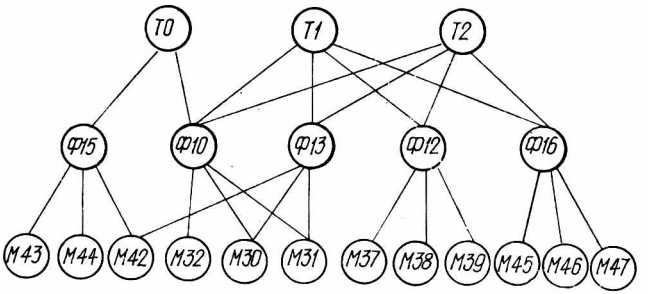

Рис. 2

С-структура имеет вид (рис. 1, 2) [3]. Итак, мы рассмотрели, какие процедуры вывода используются при разрешении личных

местоимений. Как и другие классы процедур вывода, данный класс управляется метапроцедурой, которая обслуживает весь процесс разрешения местоимений.

Список литературы: 1. *Терзиян В. Я.* Анализ, семантическая нормализация и идентификация естественноязыковых текстов.— В кн.: Интерактивные системы: Тез. докл. и сообщ. 4-й школы-семинара. Тбилиси, 1982, с. 219—221. 2. *Русская* грамматика. В 2-х т.— М.: Наука, 1980.— 784 с. 3. *Ловицкий В. А.* Диалоговая естественноязыковая система принятия решений.— Х.: Ротапринт ХПИ, 1981.— 110 с.

Поступила в редколлегию 17.11.83