

Лингвистический процессор ЭТАП:

дескрипторное соответствие и обработка метафор [1]

Л.Л. Цинман, В.Г. Сизов

ИППИ РАН

cinman@iitp.ru, sizov@iitp.ru

Синтаксический анализатор (СА) системы ЭТАП [Апресян и др. 1992] строит синтаксическую структуру (СинтС) анализируемой фразы на основе результатов морфологического анализа, осуществляемого по морфологическому словарю, и с помощью синтаксической информации, содержащейся в словарных статьях комбинаторного словаря (КС) слов анализируемой фразы. Основным инструментом получения СинтС являются синтаксические правила (синтагмы), некоторые из которых проверяют наличие (или отсутствие) в словарных статьях КС той или иной синтаксической информации.

Главной проблемой, которая возникает в любом СА при построении СинтС, является борьба с омонимией всякого рода (как морфологической, так и синтаксической). Особенно это актуально для СА системы ЭТАП. Дело в том, что в нашем СА используется фильтровый метод синтаксического анализа (СинтА). При этом методе сначала строится полный набор допустимых (с точки зрения принятого в нашем СА описания синтаксиса) гипотетических синтаксических связей между словами анализируемой фразы. Затем, с помощью разнообразных фильтров этот набор вычищается до тех пор, пока оставшиеся связи не образуют дерево, являющееся искомым СинтС фразы. [Апресян и др. 1989]. Количество гипотетических связей, первоначально возникающих при фильтровом методе СА, может в несколько раз (в зависимости от неоднозначности слов фразы и вероятной синтаксической омонимии) превосходить необходимое для СинтС число связей. Впрочем, этот недостаток фильтрового метода алгоритма с лихвой перекрывается тем фактом, что он позволяет

анализировать самые редкие и сложные синтаксические конструкции, которые только могут встретиться в тексте.

Резкого уменьшения количества возникающих при Синта гипотетических связей можно добиться за счет более полного и точного описания разных типов языковой информации в словарных статьях КС. В этой работе мы обсудим проблемы использования только одного типа такой информации – дескрипторных ограничений на актанты управляющих слов, представленных в моделях управления (МУ) этих слов.

Рассмотрим в качестве примера словарное описание глагола *резать*. Этот глагол представлен в нашем КС тремя словарными статьями. Приведем фрагменты этих словарных статей.

РЕЗАТЬ1 (Мальчик острыми ножницами режет фольгу на длинные тонкие полоски).

МУ:

Первый актант: S, им.

Второй актант: S, вин.

Третий актант: S, твор.

Четвертый актант: *на1*

TRANS:*CUT1*

РЕЗАТЬ2 (Волк острыми клыками мгновенно зарезал кабана)

МУ:

Первый актант: S, им.

Второй актант: S, вин.

Третий актант: S, твор.

TRANS:*KILL*

РЕЗАТЬ3 (Дым от костра резал мне глаза)

МУ:

Первый актант: S, им.

Второй актант: S, вин.

TRANS:*GRATE3*

Допустим теперь, что на вход СА поступила фраза

(1) *Этот станок режет любой металл.*

Три слова этой фразы имеют омонимичные разборы:

2.1 *станок* - S, им.; 2.2 *станок* - S, вин.

3.1 *резать*₁; 3.1 *резать*₂; 3.3 *резать*₃

5.1 *металл* - S, им.; 5.2 *металл* - S, вин.

Для приведенных выше фрагментов описания словарных статей глагола *резать* мы получим 12 различных вариантов синтаксического разбора этой простой фразы: 3 претендента на роль сказуемого, для каждого из которых по 2 претендента на роль подлежащего и на роль прямого дополнения. Для снятия подобной омонимии в нашем СА широко используется система семантических признаков, или дескрипторов. В словаре существительным, глаголам и некоторым прилагательным приписываются соответствующие дескрипторы (например, слово *станок* снабжено дескрипторами 'ПРЕДМЕТ', 'МЕХАНИЗМ', 'ИНСТРУМЕНТ', а слово *металл* – дескрипторами 'ВЕЩЕСТВО', 'ПРЕДМЕТ', 'МАТЕРИАЛ'). Одновременно в МУ указываются дескрипторы слов, которые могут претендовать на роль актантов ключевого слова (если, конечно, дескрипторы слов предполагаемых актантов можно перечислить).

Вот как выглядят МУ разных значений глагола *резать* с указанием дескрипторных ограничений:

РЕЗАТЬ₁ (*Мальчик острыми ножницами режет фольгу на длинные тонкие полоски*).

МУ:

Первый актант: S, им., 'ЧЕЛОВЕК', 'ИНСТРУМЕНТ', 'МЕХАНИЗМ'

Второй актант: S, вин., 'ПРЕДМЕТ'

Третий актант: S, твор., 'ИНСТРУМЕНТ', 'МЕХАНИЗМ',

Четвертый актант: *на*₁, 'ПРЕДМЕТ', 'ЧАСТЬ'

(запись дескрипторов в строке рядом с предлогом означает, что одним из этих дескрипторов должно обладать слово, управляемое этим предлогом).

РЕЗАТЬ₂ (*Волк острыми клыками мгновенно зарезал кабана*)

МУ:

Первый актант: S, им., 'ЧЕЛОВЕК', 'ЖИВОТНОЕ'

Второй актант: S, вин., 'ЧЕЛОВЕК', 'ЖИВОТНОЕ'

Третий актант: S, твор., 'ИНСТРУМЕНТ', 'ОРГАН'

РЕЗАТЬ₃ (*Дым от костра резал мне глаза*)

МУ:

Первый актант: S, им., 'ФАКТ'

Второй актант: S, вин., 'ОРГАН', 'ЧУВСТВО'

Синтаксические правила при установлении гипотетических синтаксических связей между словами проверяют требуемое дескрипторное соответствие. Связь устанавливается только в том случае, когда слову, претендующему на роль некоторого актанта, приписан хотя бы один из указанных в соответствующей строке модели управления дескриптор. При отсутствии в строке МУ каких-либо дескрипторов дескрипторное согласование предполагается выполненным.

Нетрудно убедиться в том, что при учете дескрипторных согласований фраза (1) (так же, как и фразы, использованные в качестве иллюстраций при лексемах *резать1*, *резать2* и *резать3*), получает лишь один (правильный) синтаксический разбор.

Итак, использование в МУ дескрипторных ограничений на актанты является мощным средством борьбы с омонимией на этапе СинтА. Но у широкого использования этих ограничений есть и обратная сторона. Рассмотрим еще две фразы с глаголом *резать*.

(2) *Нос корабля режет воду.*

(3) *Он при каждом удобном случае режет правду.*

В этих фразах глагол *резать* использован в нестандартных значениях: в (2) налицо метафорическое употребление *резать1*, а в (3) *резать* входит в состав фразеологической единицы. Ни в том, ни в другом случае задаваемые МУ дескрипторные соответствия не выполняются, поскольку в (2) ни для одного из значений глагола *резать* слово *нос* (ни в одном из своих значений) не может претендовать на роль подлежащего, а в (3) слово *правда* не может оказаться прямым дополнением. Поэтому наш СА не сможет правильно проанализировать эти фразы. В то же время без указания в МУ дескрипторных ограничений мы получили бы для каждой из этих фраз по несколько СинтС, среди которых были бы и правильные.

Разумеется, заранее предусмотреть все случаи неканонических употреблений предикатного слова создатели словарной статьи, не могут. Если же все-таки стремиться при описании МУ к учету всех случаев метафорического или фразеологического использования управляющего слова (Ю.Д.Апресян называет такие использования *разрешенными неправильностями*), то дескрипторные ограничения пришлось бы ослабить настолько, что их проверка стала бы бессмысленной. В реальной практике при описании МУ создатели словарных статей часто вообще избегают указания каких-либо дескрипторных ограничений не в последнюю очередь именно потому, что многие слова допускают неканоническое использование.

Для разрешения этой коллизии мы создали и включили в СА алгоритмическую процедуру, позволяющую в статьях КС писать в МУ дескрипторные ограничения с учетом только канонического использования управляющих лексем и в то же время оставляющую возможность правильного анализа фраз, в которых управляющая лексема использована неканонически.

Эта процедура достаточно проста. Дело в том, что в наших синтаксических правилах требование дескрипторного соответствия по МУ оформляется в виде соответствующего предиката, присутствующего в условиях правил. Поэтому есть возможность отследить те случаи, когда все условия правила, за исключением этого особого предиката, выполнены. Обнаружив подобный случай, мы не отказываемся от построения соответствующей гипотетической синтаксической связи, однако, помечаем ее как «условную». В дальнейшем эта гипотеза наряду с другими участвует в построении СинтС фразы, и если с ней не будет конкурировать какая-либо «безусловная» гипотеза, то она войдет в СинтС. Появление в

окончательной СинтС «условных» связей означает, как правило, использование во фразе некоторого управляющего слова в неканоническом его значении.

Эта процедура аналогична процедурам, описанным нами в [Цинман, Сизов 1999].

Литература

Апресян Ю.Д., Богуславский И.М., Иомдин Л.Л., Лазурский А.В., Перцов Н.В., Санников В.З., Цинман Л.Л. Лингвистическое обеспечение системы ЭТАП-2. М., Наука, 1989, 295 с.

Апресян Ю.Д., Богуславский И.М., Иомдин Л.Л., Лазурский А.В., Митюшин Л.Г., Санников В.З., Цинман Л.Л. Лингвистический процессор для сложных информационных систем. М., Наука, 1992, 256 с.

Цинман Л.Л., Сизов В.Г. Система Этап: процедуры ослабления синтаксических правил и их использование. Труды Международного семинара Диалог'99 по компьютерной лингвистике и ее приложениям. М., 1999, 321-325 с.

[1] Данная работа была выполнена при поддержке РФФИ (гранты № 99-06-80292 и № 99-06-80277).