**Синтаксические диаграммы КС-языков**

Синтаксические диаграммы КС-языка могут быть построены по его грамматике на основании следующих правил:

1. Для каждого нетерминала грамматики строится отдельная диаграмма, обозначенная названием этого нетерминала.

2. Нетерминалы из правых частей правил изображаются на диаграммах прямоугольниками, внутри которых записывается название нетерминала.

Терминальные символы изображаются в кружках или овалах.

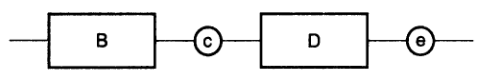
3. Для каждой правой части правила строится ветвь, представляющая собой последовательно соединенные прямоугольники и круги (овалы), следующие в том же порядке слева направо, что и соответствующие нетерминалы и терминалы правой части правила.

4. Ветви, соответствующие альтернативным правым частям правил для одного нетерминала, соединяются параллельно и образуют диаграмму для данного нетерминала.

Рассмотрим примеры построения диаграмм. Пусть в некоторой грамматике имеется правило А → а, тогда на диаграмме для нетерминала А будет ветвь

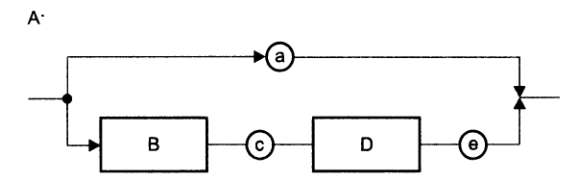


Правило А → BcDe порождает ветвь



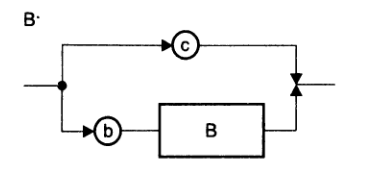
Если других правил для нетерминала А в грамматике нет, то диаграмма для этого нетерминала получается параллельным соединением ветвей. Правила для А удобней объединить в одно с альтернативными правыми частями:

А → а | BcDe.



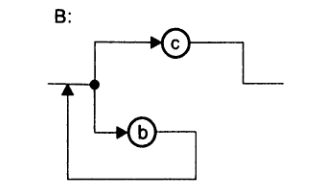
Продолжим пример. Поскольку в правилах для А фигурируют нетерминалы В и D то в грамматике должны быть правила, в которых В и D записаны в левой части. Пусть правило для В имеет вид:

В → с | bВ. Тогда строится такая диаграмма:



Можно, однако, заметить, что правила для В удовлетворяют ограничениям автоматных грамматик. А синтаксические диаграммы автоматных грамматик не должны содержать нетерминалов.

Противоречия нет. Диаграмма для В может быть преобразована. Поскольку прохождение прямоугольного блока, обозначающего B, равносильно (порождает такую же цепочку терминалов) повторному входу в диаграмму, вход в блок В можно заменить повторным входом в диаграмму.



Такое преобразование, устранившее с диаграммы нетерминальный блок B, стало возможным благодаря тому, что нетерминал В был самым правым символом в одной из альтернативных правых частей правил для В.

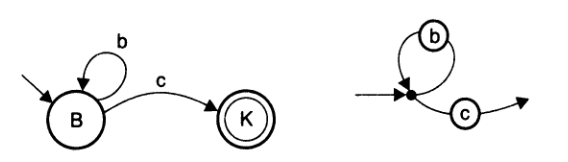
В результате преобразования концевая (правая) рекурсия заменена циклом.

Выполнив элементарное преобразование, можно нарисовать диаграмму нетерминала В в традиционном виде.

Замена правой рекурсии циклом всегда возможна (и желательна) при построении синтаксических диаграмм КС-грамматики.

Интересно заметить, что сами правила 1-4 не предусматривают циклов, в то время как на практике циклы на диаграммах имеются почти всегда.

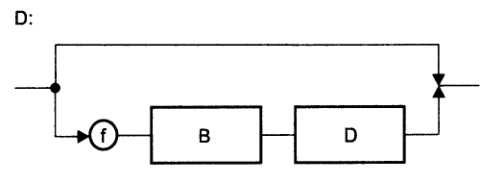
Такую же диаграмму для В можно было получить, построив фрагмент конечного автомата, а затем устранив из него состояния.



Завершая пример, зададим правило для нетерминала D:

D →fBD | ε

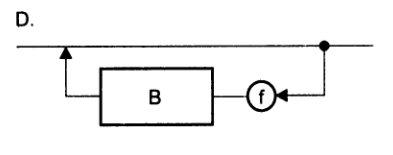
и построим диаграмму.



D:

Наличие пустой цепочки в одной из альтернативных правых частей правила приводит к появлению на диаграмме параллельной ветви, в которой нет символов.

Получившаяся диаграмма может быть, однако, снова упрощена:



**Определение языка с помощью синтаксических диаграмм**

В действительности синтаксические диаграммы, как правило, не строятся по имеющейся грамматике, а служат самостоятельным средством проектирования языков, в том числе и языков программирования.

При этом язык определяется совокупностью диаграмм, первая из которых соответствует начальному нетерминалу грамматики.

Определять синтаксис в виде совокупности диаграмм, на которых имеются нетерминальные блоки, можно не только для контекстно-свободных, но и для автоматных языков. Только из-за отсутствия самовложения диаграммы автоматного языка всегда можно объединить в одну, не содержащую нетерминалов.

Для этого достаточно «подставить» в диаграмму начального нетерминала другие диаграммы вместо соответствующих прямоугольных блоков.

Для КС-языка такая подстановка невозможна из-за самовложения.

**Язык многочленов**

Для примера построим синтаксические диаграммы, задающие правила записи (синтаксис) многочленов oт x c постоянными целочисленными коэффициентами, то есть определяющие язык многочленов.

Примеры таких многочленов:

5x3 + x2-12x+10

-х ,199

Последний пример может вызвать возражение, поскольку не содержит переменной х. Условимся, однако, и такую запись считать правильным многочленом нулевой степени.

Чтобы запись многочленов могла быть обработана компьютерной программой (транслятором или вычислителем многочленов), предусмотрим возможность записи символов «в строку» без надстрочных показателей степени.

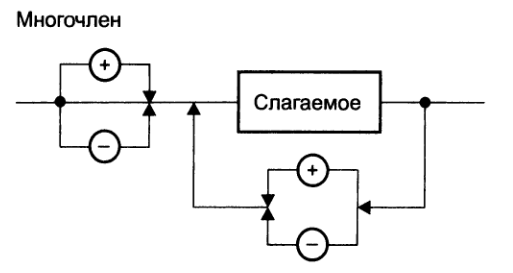
Возведение в степень будем обозначать, как это принято в языке Бейсик и некоторых диалектах Алгола, с помощью знака «^». Тогда первый пример многочлена запишется так:

5х^3 + х^2 - 12х + 10

Построим синтаксические диаграммы, определяющие правила записи многочленов. Первой будет диаграмма для начального нетерминала, который в нашем случае есть не что иное как «Многочлен».

Многочлен состоит из отдельных слагаемых, между которыми записываются знаки операций. Перед первым слагаемым также можно записать знак. Слагаемых должно быть не меньше одного.

С учетом этого получается диаграмма, показанная на рис. 17

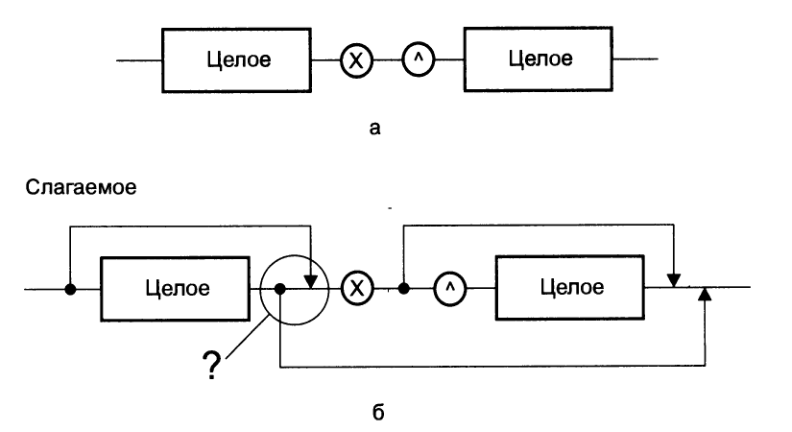


*Рис.17. Синтаксическая диаграмма многочлена*

Теперь надо построить диаграмму для нетерминала «Слагаемое», который мы ввели в грамматику многочленов.

Вначале изобразим ветвь диаграммы, соответствующую полному варианту слагаемого, когда присутствуют все его элементы: коэффициент, буква х, знак возведения в степень и сама степень (рис. 18, а).

Затем проведем «обходные» ветви, позволяющие предусмотреть такие варианты слагаемого, когда нет коэффициента (предполагается равным единице), буквы х и последующей степени, или только степени (см. рис. 18, б).

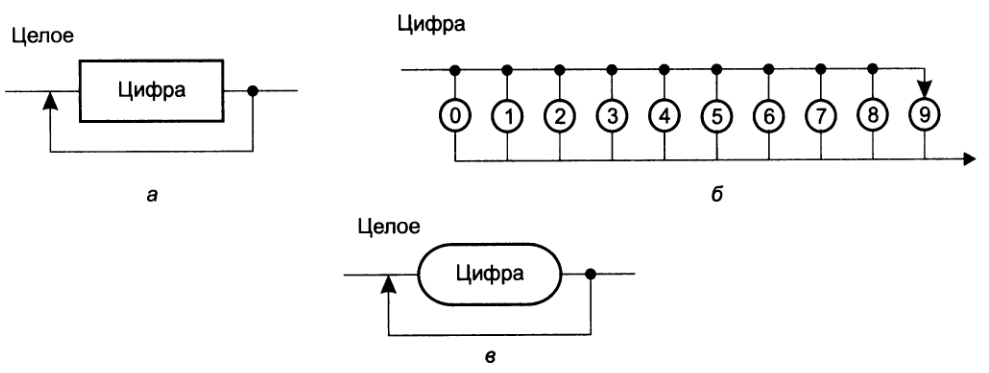


*Рис.18. Синтаксическая диаграмма слагаемого*

При этом не должно появиться такого пути на диаграмме, пройдя по которому мы минуем как коэффициент, так и х.

Коэффициент перед слагаемым и показатель степени записываются как целые числа без знака. Соответствующий нетерминал назван «Целое». В дальнейшем мы всегда будем считать (если не оговорено иное), что «целое» означает целое без знака.

Целое без знака есть последовательность, состоящая из одной или более цифр (рис. 19, а). Диаграмма для нетерминала «Цифра» показана на рис. 19, б.



*Рис. 19. Синтаксические диаграммы для целого и для цифры*

Поскольку арабские цифры используются в самых разнообразных языках, было бы неудобно каждый раз приводить диаграмму, подобную изображенной на рис. 19, б. В дальнейшем будем вместо нетерминального блока «Цифра» использовать на диаграммах овал (рис. 19, в), считая что «цифра» — это «почти терминальный символ».

Нетрудно понять (хотя бы по отсутствию самовложения), что язык многочленов — автоматный. Все диаграммы можно было бы объединить в одну, не содержащую нетерминальных блоков. Однако делать этого мы не будем.

Во-первых, несколько несложных диаграмм воспринимаются проще, чем одна громоздкая.

Во-вторых, использование промежуточных понятий, таких как, например, «Целое», позволяет избежать дублирования: нетерминал «Целое» встречается на диаграмме слагаемого дважды.

В-третьих, представив автоматный язык с помощью КС- диаграмм, мы на простом примере рассмотрим методы распознавания КС-языков, не потеряв при этом общности подхода.