Глава 4. Восходящий синтаксический анализ

4.2. Грамматики слабого предшествования

Грамматики слабого предшествования являются небольшим расширением класса грамматик простого предшествования, связанное с тем, что разрешено пересечение отношений $\lessdot$ и $\doteq$. Таким образом, в отличие от грамматик простого предшествования, где между парами символов грамматики допускается не более одного отношения предшествования, в грамматиках слабого предшествования между парами символов могут быть одновременно отношения $\lessdot$ и $\doteq$.

Для грамматик слабого предшествования отношение $\gtrdot$ по-прежнему используется для определения окончания основы. Возникают трудности с определением заголовка основы, связанные с тем, что правая часть одной продукции может быть суффиксом правой части другой продукции.

Пусть $\alpha\beta\gamma w$, $\alpha, \beta, \gamma \in (V_T \cup V_N)^*$, $w \in V_T^{\ *}$ – правосторонняя сентенциальная форма, в которой окончанием основы является последний символ строки γ. Если в грамматике есть продукции $A \to \gamma$ и $A \to \beta\gamma$, то возникает вопрос, какую из этих продукций необходимо выбрать для свертки (что будет основой, γ или $\beta\gamma$?). В этом случае между всеми символами строк β и γ выполняется отношение $\doteq$, а между последним символом строки β и первым символом строки γ выполняются отношения $\lessdot$ и $\doteq$.

В грамматиках слабого предшествования в случае такого конфликта в качестве основы для свертки выбирается наиболее длинная основа, т. е. $\beta\gamma$.

Формально КС-грамматика $G = (V_T, V_N, P, S)$ называется *грамматикой слабого предшествования*, если:

1) не содержит ε-продукций;

2) никакие две продукции грамматики не имеют совпадающих правых частей;

3) отношение $\gtrdot$ не пересекается с объединением отношений $\lessdot$ и $\doteq$;

4) для продукций $A \to \alpha X\beta$ и $B \to \beta$, где $\alpha, \beta \in (V_T \cup V_N)^*$, $X \in V_T \cup V_N$, не выполняется ни отношение $X \lessdot B$, ни отношение $X \doteq B$.

Для поиска окончания основы (как и для грамматик простого предшествования) достаточно проанализировать сентенциальную форму слева направо и найти самую левую пару символов X_j и X_{j+1}, таких, что $X_j \gtrdot X_{j+1}$, т. е. X_j – окончание основы. Затем сентенциальная форма просматривается справа налево, начиная с символа X_j, до тех пор, пока не будет найдена пара символов X_{i-1} и X_i, таких, что $X_{i-1} \lessdot X_i$ и не выполняется отношение $X_{i-1} \doteq X_i$, т. е. X_i – заголовок основы. Ясно, что между всеми соседними символами внутри основы выполняется либо отношение $\doteq$, либо отношения $\lessdot$ и $\doteq$. Таким образом, в процессе поиска заголовка основы при просмотре сентенциальной формы справа налево в случае, если между парой соседних символов выполняется как отношение $\lessdot$, так и отношение $\doteq$, то отношение $\doteq$ имеет приоритет над отношением $\lessdot$.

Детали реализации алгоритма синтаксического анализа предлагаются в качестве упражнения.

4.2. Грамматики операторного предшествования

Грамматики операторного предшествования представляют достаточно широкий класс КС-грамматик.

Операторной грамматикой называется ε-свободная (не содержит ε-продукций) КС-грамматика, в которой правые части всех продукций не содержат смежных нетерминалов. В таких грамматиках терминалы можно рассматривать как операции, а нетерминалы – как операнды. Например, в арифметических выражениях можно сказать, что операция умножения *предшествует* операции сложения, поскольку умножение имеет более высокий приоритет, чем сложение. Порядок вычисления значения арифметического выражения определяется только порядком выполнения операций и не зависит от операндов. Поэтому понятие предшествования можно определить только для операций (терминалов).

Пусть $G = (V_T, V_N, P, S)$ – операторная КС-грамматика, пополненная продукцией $S' \to \perp S \perp$, где $\perp \in V_T$. *Отношения операторного предшествования* задаются на множестве $V_T \times V_T$ следующим образом:

1. $a \doteq b$, если существует некоторая продукция $A \to \alpha a C b \beta$, $A \in V_N$, $\alpha, \beta \in (V_T \cup V_N)^*$, $C \in V_N \cup \{\varepsilon\}$.

2. $a \lessdot b$, если существует некоторая продукция $A \to \alpha a B \beta$, $A, B \in V_N$, $\alpha, \beta \in (V_T \cup V_N)^*$, такая, что $B \overset{+}{\Rightarrow} Cb\delta$, $C \in V_N \cup \{\varepsilon\}$, $\delta \in (V_T \cup V_N)^*$.

3. $\perp \lessdot a$, если $S \overset{+}{\Rightarrow} Ca\alpha$, $C \in V_N \cup \{\varepsilon\}$, $\alpha \in (V_T \cup V_N)^*$.

4. $a \gtrdot b$, если существует некоторая продукция $A \to \alpha B b \beta$, $A, B \in V_N$, $\alpha, \beta \in (V_T \cup V_N)^*$, такая, что $B \overset{+}{\Rightarrow} \delta a C$, $C \in V_N \cup \{\varepsilon\}$, $\delta \in (V_T \cup V_N)^*$.

5. $a \gtrdot \perp$, если $S \overset{+}{\Rightarrow} \alpha a C$, $C \in V_N \cup \{\varepsilon\}$, $\alpha \in (V_T \cup V_N)^*$.

Отношения операторного предшествования можно задавать с помощью *матрицы операторного предшествования*. Строки и столбцы матрицы соответствуют символам из V_T. Пустой элемент матрицы соответствует синтаксической ошибке, например, если для пары a и b элемент пустой, то ни в одной правильной входной строке b не может следовать непосредственно за a.

Операторная КС-грамматика $G = (V_T, V_N, P, S)$ называется *грамматикой операторного предшествования*, если:

1) никакие две продукции грамматики не имеют совпадающих правых частей;

2) между любыми двумя терминалами из множества $V_T \times V_T$ выполняется не более одного отношения операторного предшествования.

Отношения операторного предшествования вычисляются так же, как и для грамматик простого предшествования, определив для каждого нетерминала X множества $L(X)$ и $R(X)$. Отличие заключается в том, что в эти множества включаются только терминалы, игнорируя нетерминалы (подставляя вместо нетерминалов пустую строку ε).

Построим матрицу операторного предшествования для грамматики (в предположении, что имеется продукция $S \to \perp E \perp$)

$E \to E + T \mid T$

$T \to T \times F \mid F$

$F \to (E) \mid i$

Определим множества $L(X)$ и $R(X)$ для $X \in \{E, T, F\}$:

$L(E) = \{E, T, F, (, i\} = \{+, \times, (, i\}$;

$L(T) = \{T, F, (, i\} = \{\times, (, i\}$;

$L(F) = \{(, i\}$;

$R(E) = \{T, F,), i\} = \{+, \times,), i\}$;

$R(T) = \{F,), i\} = \{\times,), i\}$;

$R(F) = \{), i\}$;

Соответствующая матрица операторного предшествования приведена на рис. 4.9.

	$+$	$\times$	(	)	i	$\perp$
$+$	$\gtrdot$	$\lessdot$	$\lessdot$	$\gtrdot$	$\lessdot$	$\gtrdot$
$\times$	$\gtrdot$	$\gtrdot$	$\lessdot$	$\gtrdot$	$\lessdot$	$\gtrdot$
(	$\lessdot$	$\lessdot$	$\lessdot$	$\doteq$	$\lessdot$	
)	$\gtrdot$	$\gtrdot$		$\gtrdot$		$\gtrdot$
i	$\gtrdot$	$\gtrdot$		$\gtrdot$		$\gtrdot$
$\perp$	$\lessdot$	$\lessdot$	$\lessdot$		$\lessdot$	

Рис. 4.9. Матрица операторного предшествования

При распознавании основы возникает проблема с нетерминалами, поскольку для них не определены отношения операторного предшествования. Чтобы решить эту проблему исходная грамматика преобразуется в так называемую остовную грамматику путем замены всех нетерминалов одним начальным нетерминалом и устранением всех цепных продукций.

Пусть $G = (V_T, V_N, P, S)$ – операторная грамматика. *Остовной грамматикой* для грамматики G называется грамматика $G_S = (V_T, \{S\}, P', S)$, множество продукций P' которой строится следующим образом:

а) если множество P грамматики G содержит продукцию вида $A \to Y_1Y_2...Y_n$, где $Y_i \in V_T \cup V_N$, $1 \le i \le n$, то в P' включается продукция $S \to X_1X_2...X_n$, где $X_i = Y_i$, если $Y_i \in V_T$, или $X_i = S$, если $Y_i \in V_N$;

б) множество продукций P' не должно содержать продукций вида $S \to S$.

Например, для рассмотренной выше грамматики арифметических выражений остовной будет грамматика с продукциями

$E \to E + E \mid E \times E \mid (E) \mid i$

Следует заметить, что язык $L(G) \subseteq L(G_S)$ и грамматика G_S может порождать строки, не принадлежащие $L(G)$. Кроме того, грамматика G_S может быть неоднозначной. Однако отношения операторного предшествования гарантируют единственность синтаксического разбора и его правильность.

Поиск основы реализуется также как и для грамматик простого предшествования. Отличие заключается в том, что если в вершине стека оказывается нетерминал, то для определения отношения предшествования рассматривается ближайший к вершине стека терминал (нетерминал игнорируется). Детали реализации алгоритма синтаксического анализа предлагаются в качестве упражнения.

Рассмотрим работу анализатора на примере разбора строки $i \times (i + i)\perp$, которая выводится в соответствии со следующей правосторонней схемой (символ $\perp$ начала строки опущен)

$$E\perp \Rightarrow E \times E\perp \Rightarrow E \times (E)\perp \Rightarrow E \times (E + E)\perp \Rightarrow E \times (E + i)\perp \Rightarrow$$
$$\Rightarrow E \times (i + i)\perp \Rightarrow i \times (i + i)\perp.$$

Процесс разбора показан на рис. 4.10.

	$+$	$\times$	$($	$)$	i	$\perp$
$+$	$\gtrdot$	$\lessdot$	$\lessdot$	$\gtrdot$	$\lessdot$	$\gtrdot$
$\times$	$\gtrdot$	$\gtrdot$	$\lessdot$	$\gtrdot$	$\lessdot$	$\gtrdot$
$($	$\lessdot$	$\lessdot$	$\lessdot$	$\doteq$	$\lessdot$	
$)$	$\gtrdot$	$\gtrdot$		$\gtrdot$		$\gtrdot$
i	$\gtrdot$	$\gtrdot$		$\gtrdot$		$\gtrdot$
$\perp$	$\lessdot$	$\lessdot$	$\lessdot$		$\lessdot$	

Входной буфер	Содержимое стека	Основа	Выполняемое действие
$i\times(i+i)\perp$	$\perp$		Перенос i в стек, т. к. $\perp \lessdot i$
$\times(i+i)\perp$	$\perp i$	i	Свертка для $E \to i$, т. к. $i \gtrdot \times$
$\times(i+i)\perp$	$\perp E$		Перенос $\times$ в стек, т. к. $\perp \lessdot \times$
$(i+i)\perp$	$\perp E\times$		Перенос (в стек, т. к. $\times \lessdot ($
$i+i)\perp$	$\perp E\times($		Перенос i в стек, т. к. $(\lessdot i$
$+i)\perp$	$\perp E\times(i$	i	Свертка для $E \to i$, т. к. $i \gtrdot +$
$+i)\perp$	$\perp E\times(E$		Перенос + в стек, т. к. $(\lessdot +$
$i)\perp$	$\perp E\times(E+$		Перенос i в стек, т. к. $+ \lessdot i$
$)\perp$	$\perp E\times(E+i$	i	Свертка для $E \to i$, т. к. $i \gtrdot)$
$)\perp$	$\perp E\times(E+E$	$E+E$	Свертка для $E \to E+E$, т. к. $+ \gtrdot)$
$)\perp$	$\perp E\times(E$		Перенос) в стек, т. к. $(\doteq)$
$\perp$	$\perp E\times(E)$	(E)	Свертка для $E \to (E)$, т. к. $) \gtrdot \perp$
$\perp$	$\perp E\times E$	$E\times E$	Свертка для $E \to E\times E$, т. к. $\times \gtrdot \perp$
$\perp$	$\perp E$		Разбор успешно завершен

Рис. 4.10. Процесс разбора строки $i \times (i + i)\perp$