

# Решение задачи по построению порождающей грамматики

## 1. Построение грамматики по описанию языка

Исходное описание языка:

$\langle S \rangle ::= \text{'if' } [ \langle E \rangle ] ( \text{'i' } \text{' : ' } \text{'then' } \langle O \rangle ) \dots$   
 $\langle E \rangle ::= \text{'i' } \mid \text{'i' } \text{' < > ' } \langle E \rangle$   
 $\langle O \rangle ::= \text{'o' } \langle O \rangle \mid \langle S \rangle \mid \text{'o'}$

Интерпретация:

- [...] - опциональный элемент (0 или 1 вхождение)
- (...)... - повторение 1 или более раз

Построенная контекстно-свободная грамматика  $G = (V, \Sigma, P, S)$ :

$$\begin{aligned} S &\rightarrow \text{if } E \text{ } B \mid \text{if } B \\ E &\rightarrow i \mid i \text{ } \langle \rangle E \\ O &\rightarrow o \text{ } O \mid S \mid o \\ B &\rightarrow i : \text{ then } O \text{ } B \mid i : \text{ then } O \end{aligned}$$

Где:

- $V = \{S, E, O, B\}$  - множество нетерминалов
- $\Sigma = \{\text{if}, i, \langle \rangle, :, \text{then}, o\}$  - множество терминалов
- $S$  - начальный символ

## 2. Тип грамматики и свойства

Построенная грамматика является:

### 1. Контекстно-свободной (тип 2 по иерархии Хомского):

- Все правила имеют форму  $A \rightarrow \alpha$ , где  $A \in V$ ,  $\alpha \in (V \cup \Sigma)^*$
- Левая часть каждого правила - одиночный нетерминал

### 2. Рекурсивной:

- Прямая рекурсия:  $O \rightarrow o O$
- Косвенная рекурсия:  $O \rightarrow S, S \rightarrow \dots \rightarrow O$

### 3. Неоднозначной:

- Для строки "o" есть два вывода:  $O \Rightarrow o$  и  $O \Rightarrow o O \Rightarrow o$

### 4. С левой рекурсией:

- В правиле  $O \rightarrow o O$

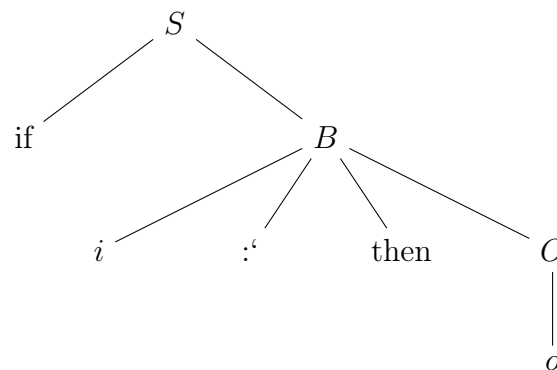
### 5. Не левофакторизованная:

- У правила  $O$  есть две альтернативы, начинающиеся с  $o$ :  $O \rightarrow o O$  и  $O \rightarrow o$

## 3. Деревья разбора для трёх примеров

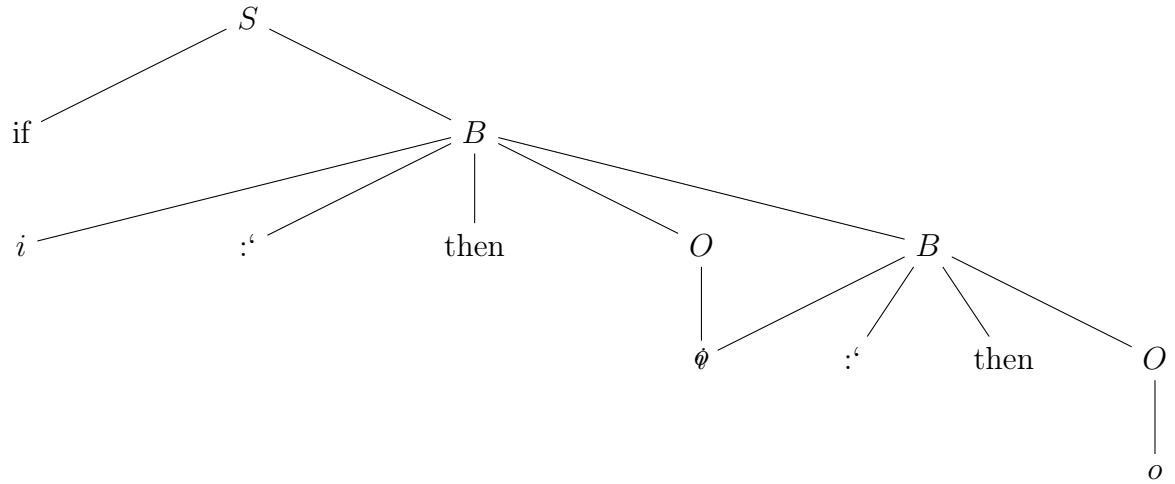
### Пример 1: if i : then o

$S \Rightarrow \text{if } B \Rightarrow \text{if } (i : \text{then } O) \Rightarrow \text{if } i : \text{then } o$



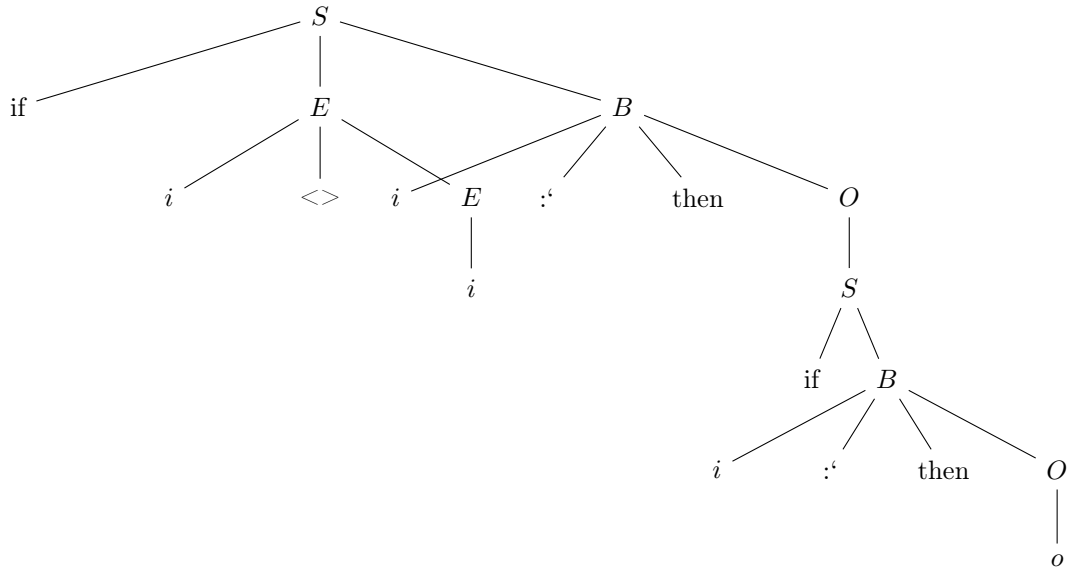
### Пример 2: if i i : then o i : then o

$S \Rightarrow \text{if } B$   
 $\Rightarrow \text{if } (i : \text{then } O B)$   
 $\Rightarrow \text{if } i : \text{then } o (i : \text{then } O)$   
 $\Rightarrow \text{if } i : \text{then } o i : \text{then } o$



**Пример 3:** if i <> i i : then if i : then o

$S \Rightarrow \text{if } E \ B$   
 $\Rightarrow \text{if } (i \ <> \ E) \ B$   
 $\Rightarrow \text{if } i \ <> \ i \ (i : \text{ then } O)$   
 $\Rightarrow \text{if } i \ <> \ i \ i : \text{ then } S$   
 $\Rightarrow \text{if } i \ <> \ i \ i : \text{ then if } B$   
 $\Rightarrow \text{if } i \ <> \ i \ i : \text{ then if } (i : \text{ then } O)$   
 $\Rightarrow \text{if } i \ <> \ i \ i : \text{ then if } i : \text{ then } o$



#### 4. Устранение левой рекурсии

Грамматика содержит прямую левую рекурсию в правиле:

$$O \rightarrow o O \mid S \mid o$$

Устраним левую рекурсию. Введём новый нетерминал  $O'$ :

$$\begin{aligned} O &\rightarrow S O' \mid o O' \\ O' &\rightarrow O O' \mid \varepsilon \end{aligned}$$

Но это не совсем корректно, так как получается правая рекурсия. Правильнее:

$$\begin{aligned} O &\rightarrow S O' \mid o O' \\ O' &\rightarrow \varepsilon \mid O O' \end{aligned}$$

Это всё ещё содержит рекурсию. Используем стандартный алгоритм устранения левой рекурсии:

Исходное правило:  $O \rightarrow o O \mid S \mid o$

1. Выносим общий префикс для альтернатив, начинающихся с  $o$ :

$$O \rightarrow o O \mid o \mid S$$

2. Факторизуем:

$$O \rightarrow o (O \mid \varepsilon) \mid S$$

3. Вводим новый нетерминал  $X$ :

$$\begin{aligned} O &\rightarrow o X \mid S \\ X &\rightarrow O \mid \varepsilon \end{aligned}$$

Но здесь  $X \rightarrow O$  создаёт левую рекурсию. Правильный подход:

$$\begin{aligned} O &\rightarrow S \mid o Y \\ Y &\rightarrow O \mid \varepsilon \end{aligned}$$

Заметим, что  $Y \rightarrow O$  даёт  $O \rightarrow o Y \rightarrow o O$ , что эквивалентно исходному. Но это уже не левая рекурсия, а правая. Фактически мы преобразовали левую рекурсию в правую.

## 5. Левофакторизация грамматики

Грамматика не является левофакторизованной, так как:

- У правила  $S$  есть две альтернативы, начинающиеся с "if"
- У правила  $O$  есть две альтернативы, начинающиеся с  $o$

Проведём левофакторизацию:

Для правила  $S$ :

$$S \rightarrow \text{if } E B \mid \text{if } B$$

Выносим общий префикс "if":

$$\begin{aligned} S &\rightarrow \text{if } Z \\ Z &\rightarrow E B \mid B \end{aligned}$$

Для правила  $O$ :

После устранения левой рекурсии получаем:

$$\begin{aligned} O &\rightarrow S \mid o Y \\ Y &\rightarrow O \mid \varepsilon \end{aligned}$$

Это правило уже левофакторизовано.

Для правила  $B$ :

$$B \rightarrow i : \text{then } O \ B \mid i : \text{then } O$$

Выносим общий префикс " $i : \text{then } O$ ":

$$\begin{aligned} B &\rightarrow i : \text{then } O \ W \\ W &\rightarrow B \mid \varepsilon \end{aligned}$$

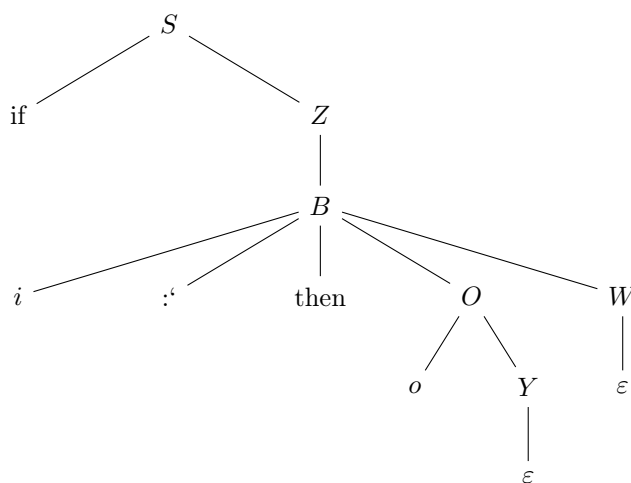
## 6. Модифицированная грамматика

После устранения левой рекурсии и левофакторизации получаем модифицированную грамматику  $G'$ :

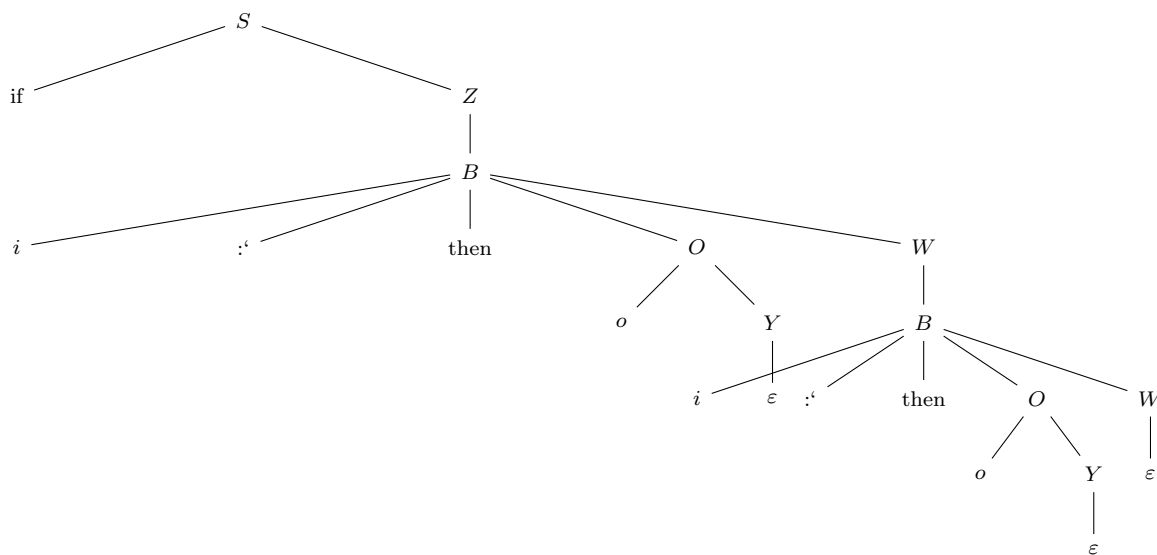
$S$	$\rightarrow$	$\text{if } Z$
$Z$	$\rightarrow$	$E \ B \mid B$
$E$	$\rightarrow$	$i \ F$
$F$	$\rightarrow$	$<> \ E \mid \varepsilon$
$B$	$\rightarrow$	$i : \text{then } O \ W$
$W$	$\rightarrow$	$B \mid \varepsilon$
$O$	$\rightarrow$	$S \mid o \ Y$
$Y$	$\rightarrow$	$O \mid \varepsilon$

## 7. Деревья разбора для модифицированной грамматики

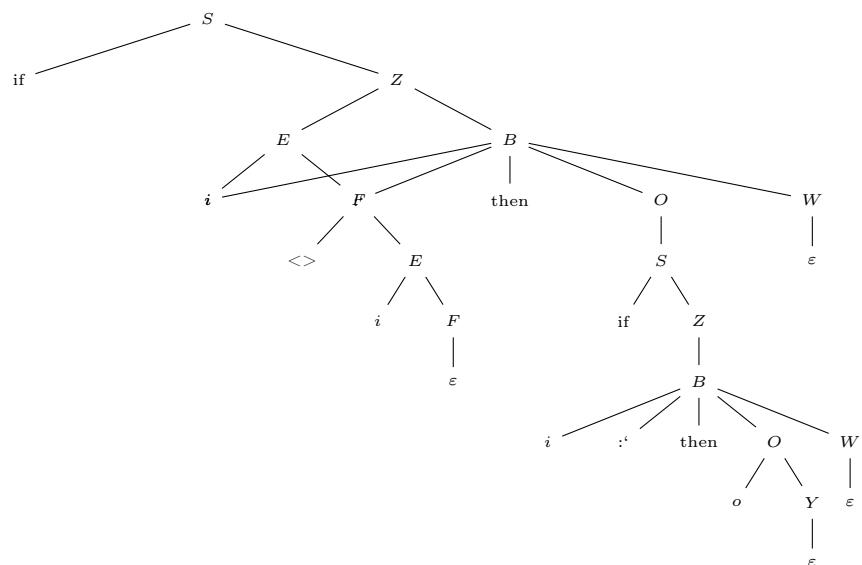
Пример 1: if i : then o



Пример 2: if i i : then o i : then o



Пример 3: if i <> i i : then if i : then o



## 8. Сравнение грамматик

- **Исходная грамматика:** более компактная и понятная, но содержит левую рекурсию и не левофакторизована
- **Модифицированная грамматика:** не содержит левой рекурсии, левофакторизована, пригодна для нисходящего синтаксического анализа, но более сложная
- **Эквивалентность:** обе грамматики порождают один и тот же язык
- **Деревья разбора:** в модифицированной грамматике деревья более глубокие из-за введения дополнительных нетерминалов