$S \overset{*}{\Rightarrow} \alpha A_{n_1}\beta \overset{*}{\Rightarrow} \alpha\omega_1\beta; S \overset{*}{\Rightarrow} \alpha\gamma A_{n_2}\delta\beta \overset{*}{\Rightarrow} \alpha\gamma\omega_2\delta\beta$, при этом ω_2 является подцепочкой ω_1.

Заменим в изначальном дереве узел n_1 на n_2. Полученное дерево является деревом вывода $\alpha\omega_2\delta$.

Повторяем процесс замены одинаковых нетерминалов до тех пор, пока в дереве не останутся только уникальные нетерминалы.

В полученном дереве не может быть ветвей длины большей, чем m.

По построению оно является деревом вывода. □

4.3 Нормальная форма Хомского

Определение 4.3.1. Контекстно-свободная грамматика $\langle \Sigma, N, P, S \rangle$ находится в *Нормальной Форме Хомского*, если она содержит только правила следующего вида:

- $A \to BC$, где $A, B, C \in N$, S не содержится в правой части правила
- $A \to a$, где $A \in N, a \in \Sigma$
- $S \to \varepsilon$

Теорема 4.3.1. Любую КС грамматику можно преобразовать в НФХ.

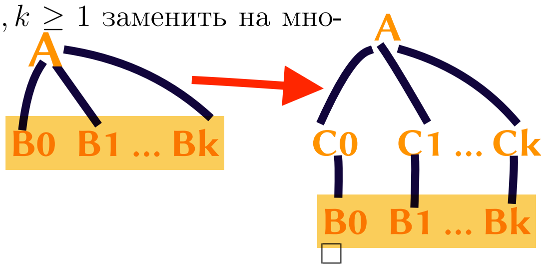
Доказательство. Алгоритм преобразования в НФХ состоит из следующих шагов:

- Замена неодиночных терминалов
- Удаление длинных правил
- Удаление ε-правил
- Удаление цепных правил
- Удаление бесполезных нетерминалов

То, что каждый из этих шагов преобразует грамматику к эквивалентной, при этом является алгоритмом, доказано в следующих леммах. □

Лемма 4.3.2. Для любой КС-грамматики можно построить эквивалентную, которая не содержит правила с неодиночными терминалами.

Доказательство. Каждое правило $A \to B_0B_1 \dots B_k, k \geq 1$ заменить на множество правил:

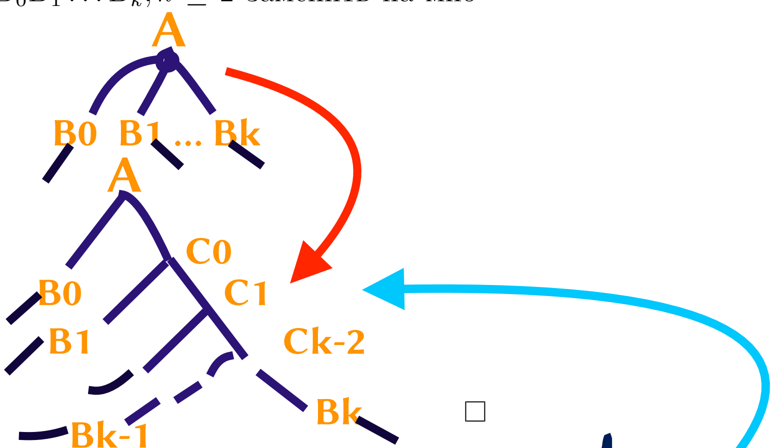


- $A \to C_0C_1 \dots C_k$
- $\{C_i \to B_i \mid B_i \in \Sigma, C_i$ — новый нетерминал$\}$

□

Лемма 4.3.3. Для любой КС-грамматики можно построить эквивалентную, которая не содержит правил длины больше 2.

Доказательство. Каждое правило $A \to B_0B_1 \dots B_k, k \geq 2$ заменить на множество правил:



- $A \to B_0C_0$
- $C_0 \to B_1C_1$
- ...
- $C_{k-3} \to B_{k-2}C_{k-2}$
- $C_{k-2} \to B_{k-1}B_k$

□

Лемма 4.3.4. Для любой КС-грамматики можно построить эквивалентную, не содержащую ε-правил.

Доказательство. Определим ε-правила:

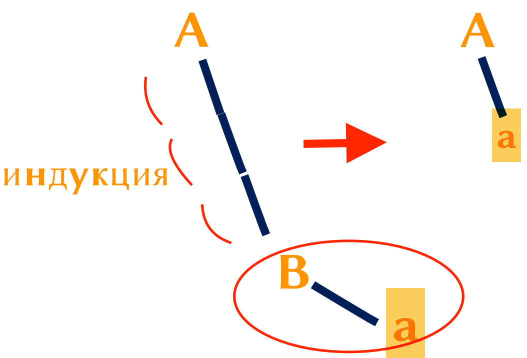
- $A \to \varepsilon$
- $A \to B_0 \dots B_k, \forall i: \ B_i$ — ε-правило.

Каждое правило $A \to B_0B_1 \dots B_k$ заменяем на множество правил, где каждое ε-правило удалено во всех возможных комбинациях. □

Лемма 4.3.5. Можно удалить все цепные правила

Доказательство. *Цепное правило* — правило вида $A \to B$, где $A, B \in N$. *Цепная пара* — упорядоченная пара (A, B), в которой $A \overset{*}{\Rightarrow} B$, используя только цепные правила.

Алгоритм:



1. Найти все цепные пары в грамматике G. Найти все цепные пары можно по индукции: Базис: (A, A) — цепная пара для любого нетерминала, так как $A \overset{*}{\Rightarrow} A$ за ноль шагов. Индукция: Если пара (A, B_0) — цепная, и есть правило $B_0 \to B_1$, то (A, B_1) — цепная пара.

2. Для каждой цепной пары (A, B) добавить в грамматику G' все правила вида $A \to a$, где $B \to a$ — нецепное правило из G.

3. Удалить все цепные правила

Пусть G — контекстно-свободная грамматика. G' — грамматика, полученная в результате применения алгоритма к G. Тогда $L(G) = L(G')$. □

Определение 4.3.2. Нетерминал A называется *порождающим*, если из него может быть выведена конечная терминальная цепочка. Иначе он называется *непорождающим*.

Лемма 4.3.6. Можно удалить все бесполезные (непорождающие) нетерминалы

Доказательство. После удаления из грамматики правил, содержащих непорождающие нетерминалы, язык не изменится, так как непорождающие нетерминалы по определению не могли участвовать в выводе какого-либо слова.

Алгоритм нахождения порождающих нетерминалов:

1. Множество порождающих нетерминалов пустое.

2. Найти правила, не содержащие нетерминалов в правых частях и добавить нетерминалы, встречающихся в левых частях таких правил, в множество.

3. Если найдено такое правило, что все нетерминалы, стоящие в его правой части, уже входят в множество, то добавить в множество нетерминалы, стоящие в его левой части.

4. Повторить предыдущий шаг, если множество порождающих нетерминалов изменилось.

В результате получаем множество всех порождающих нетерминалов грамматики, а все нетерминалы, не попавшие в него, являются непорождающими. Их можно удалить. □

Пример 4.3.1. Приведем в Нормальную Форму Хомского однозначную грамматику правильных скобочных последовательностей: $S \to aSbS \mid \varepsilon$

Первым шагом добавим новый нетерминал и сделаем его стартовым:

$$
\begin{aligned}
S_0 &\to S \\
S &\to aSbS \mid \varepsilon
\end{aligned}
$$

Заменим все терминалы на новые нетерминалы:

$$
\begin{aligned}
S_0 &\to S \\
S &\to LSRS \mid \varepsilon \\
L &\to a \\
R &\to b
\end{aligned}
$$

Избавимся от длинных правил:

$$
\begin{aligned}
S_0 &\to S \\
S &\to LS' \mid \varepsilon \\
S' &\to SS'' \\
S'' &\to RS \\
L &\to a \\
R &\to b
\end{aligned}
$$

Избавимся от ε-продукций:

$$
\begin{aligned}
S_0 &\to S \mid \varepsilon \\
S &\to LS' \\
S' &\to S'' \mid SS'' \\
S'' &\to R \mid RS \\
L &\to a \\
R &\to b
\end{aligned}
$$

Избавимся от цепных правил:

$$\begin{aligned}
S_0 &\to LS' \mid \varepsilon \\
S &\to LS' \\
S' &\to b \mid RS \mid SS'' \\
S'' &\to b \mid RS \\
L &\to a \\
R &\to b
\end{aligned}$$

Определение 4.3.3. Контекстно-свободная грамматика $\langle \Sigma, N, P, S \rangle$ находится в *ослабленной Нормальной Форме Хомского*, если она содержит только правила следующего вида:

- $A \to BC$, где $A, B, C \in N$
- $A \to a$, где $A \in N, a \in \Sigma$
- $A \to \varepsilon$, где $A \in N$

То есть ослабленная НФХ отличается от НФХ тем, что:

1. ε может выводиться из любого нетерминала
2. S может появляться в правых частях правил