Лекция А8
Нормальная форма Хомского

Вадим Пузаренко

23 октября 2023 г.

Основные понятия

Определение А8.1.

Говорят, что КС-грамматика $\mathfrak{G} = (V, \Sigma, P, S)$ находится в **нормальной форме Хомского (НФХ)**, если её продукции имеют вид $A \longrightarrow BC$, $A \longrightarrow a$, где $A, B, C \in V$, $a \in \Sigma$.

Основные понятия

Определение А8.1.

Говорят, что КС-грамматика $\mathfrak{G} = (V, \Sigma, P, S)$ находится в **нормальной форме Хомского (НФХ)**, если её продукции имеют вид $A \longrightarrow BC$, $A \longrightarrow a$, где $A, B, C \in V$, $a \in \Sigma$.

Основная цель.

Для любого непустого КС-языка, не содержащего ε, построить грамматику, находящуюся в НФХ.

Основные понятия

Определение А8.1.

Говорят, что КС-грамматика $\mathfrak{G} = (V, \Sigma, P, S)$ находится в **нормальной форме Хомского (НФХ)**, если её продукции имеют вид $A \longrightarrow BC$, $A \longrightarrow a$, где $A, B, C \in V$, $a \in \Sigma$.

Основная цель.

Для любого непустого КС-языка, не содержащего ε, построить грамматику, находящуюся в НФХ.

Алгоритмы.

1. Удалить бесполезные символы.
2. Удалить ε-продукции ($A \longrightarrow \varepsilon$).
3. Удалить цепные продукции ($A \longrightarrow B$).

Бесполезные символы

Определение А8.2.

Символ $X \in V \cup \Sigma$ называется **полезным** в грамматике $\mathfrak{G} = (V, \Sigma, P, S)$, если существует некоторое порождение $S \Rightarrow^* \alpha\hat{}X\hat{}\beta \Rightarrow^* \gamma$, где $\alpha, \beta, \gamma \in \Sigma^*$. Символ X называется **бесполезным**, если он не является полезным.

Бесполезные символы

Определение A8.2.

Символ $X \in V \cup \Sigma$ называется **полезным** в грамматике $\mathfrak{G} = (V, \Sigma, P, S)$, если существует некоторое порождение $S \Rightarrow^* \alpha\hat{\ }X\hat{\ }\beta \Rightarrow^* \gamma$, где $\alpha, \beta, \gamma \in \Sigma^*$. Символ X называется **бесполезным**, если он не является полезным.

Свойства полезных символов.

1. Символ X называется **порождающим**, если $X \Rightarrow^* \alpha$ для некоторого $\alpha \in \Sigma^*$. Заметим, что $X \in \Sigma \cup \{\varepsilon\}$ — порождающий символ.
2. Символ X называется **достижимым**, если $S \Rightarrow^* \alpha\hat{\ }X\hat{\ }\beta$ для некоторых $\alpha, \beta \in \Sigma^*$.

Любой полезный символ является одновременно и порождающим, и достижимым.

Удаление бесполезных символов

Теорема А8.1.

Пусть $\mathfrak{G} = (V, \Sigma, P, S)$ такова, что $L(\mathfrak{G}) \neq \varnothing$, и пусть $\mathfrak{G}_1 = (V_1, \Sigma_1, P_1, S)$ — грамматика, полученная с помощью следующих двух шагов:

1. вначале удаляются все непорождающие символы и продукции, их содержащие (в результате получим грамматику $\mathfrak{G}_2 = (V_2, \Sigma_2, P_2, S)$);
2. затем удаляются все символы, не достижимые в $\mathfrak{G}_2$.

Тогда $\mathfrak{G}_1$ не имеет бесполезных символов и $L(\mathfrak{G}) = L(\mathfrak{G}_1)$.

Удаление бесполезных символов

Теорема A8.1.

Пусть $\mathfrak{G} = (V, \Sigma, P, S)$ такова, что $L(\mathfrak{G}) \neq \varnothing$, и пусть $\mathfrak{G}_1 = (V_1, \Sigma_1, P_1, S)$ — грамматика, полученная с помощью следующих двух шагов:

1. вначале удаляются все непорождающие символы и продукции, их содержащие (в результате получим грамматику $\mathfrak{G}_2 = (V_2, \Sigma_2, P_2, S)$);
2. затем удаляются все символы, не достижимые в $\mathfrak{G}_2$.

Тогда $\mathfrak{G}_1$ не имеет бесполезных символов и $L(\mathfrak{G}) = L(\mathfrak{G}_1)$.

Доказательство.

Пусть $X \in V_1 \cup \Sigma_1$; тогда $X \Rightarrow^*_{\mathfrak{G}} \alpha$ для некоторого $\alpha \in \Sigma^*$. Кроме того, каждый символ, использованный в порождении α из X, также является порождающим. Таким образом, $X \Rightarrow^*_{\mathfrak{G}_2} \alpha$.

Удаление бесполезных символов

Доказательство (продолжение).

Так как X не был удален после второго шага, найдутся $\beta \in \Sigma^*$ и $\gamma \in \Sigma^*$, для которых $S \Rightarrow^*_{\mathfrak{G}_2} \beta\hat{\ }X\hat{\ }\gamma$. Кроме того, каждый символ, использованный в этом порождении, достижим, поэтому $S \Rightarrow^*_{\mathfrak{G}_1} \beta\hat{\ }X\hat{\ }\gamma$.
Известно, что каждый символ в цепочке $\beta\hat{\ }X\hat{\ }\gamma$ достижим, и что все эти символы принадлежат $V_2 \cup \Sigma_2$, поэтому каждый из них является порождающим в $\mathfrak{G}_2$. Порождение терминальной цепочки $\beta\hat{\ }X\hat{\ }\gamma \Rightarrow^*_{\mathfrak{G}_2} \beta\hat{\ }\alpha\hat{\ }\gamma$ содержит только символы, достижимые из S, поскольку они достижимы из символов цепочки $\beta\hat{\ }X\hat{\ }\gamma$. Таким образом, это порождение есть также порождение в $\mathfrak{G}_1$, т.е. $S \Rightarrow^*_{\mathfrak{G}_1} \beta\hat{\ }X\hat{\ }\gamma \Rightarrow^*_{\mathfrak{G}_1} \beta\hat{\ }\alpha\hat{\ }\gamma$.
Итак, X полезен в $\mathfrak{G}_1$. Ввиду произвольности X в $\mathfrak{G}_1$, заключаем, что $\mathfrak{G}_1$ не содержит бесполезных символов.

Удаление бесполезных символов

Доказательство (окончание).

Докажем теперь, что $L(\mathfrak{G}_1) = L(\mathfrak{G})$.
($\subseteq$) Очевидно, поскольку все символы и продукции $\mathfrak{G}_1$ входят и в $\mathfrak{G}$.
($\supseteq$) Пусть $\alpha \in L(\mathfrak{G})$; тогда $S \Rightarrow^*_{\mathfrak{G}} \alpha$. Следовательно, каждый символ в этом порождении является порождающим, поэтому $S \Rightarrow^*_{\mathfrak{G}_2} \alpha$. Кроме того, все символы данного порождения являются достижимыми в $\mathfrak{G}_2$ и, следовательно, $S \Rightarrow^*_{\mathfrak{G}_1} \alpha$; таким образом, $\alpha \in L(\mathfrak{G}_1)$. $\square$

Алгоритм нахождения порождающих

Пусть $\mathfrak{G} = (V, \Sigma, P, S)$ — грамматика.
Базис. Если $a \in \Sigma$, то a — порождающий.
Индукция. Если $A \longrightarrow \alpha$ и цепочка α состоит только из порождающих (возможно, $\alpha = \varepsilon$), то A — порождающий.

Алгоритм нахождения порождающих

Пусть $\mathfrak{G} = (V, \Sigma, P, S)$ — грамматика.
Базис. Если $a \in \Sigma$, то a — порождающий.
Индукция. Если $A \longrightarrow \alpha$ и цепочка α состоит только из порождающих (возможно, $\alpha = \varepsilon$), то A — порождающий.

Теорема А8.2.

Данный алгоритм находит в точности все порождающие грамматики $\mathfrak{G}$.

Алгоритм нахождения порождающих

Пусть $\mathfrak{G} = (V, \Sigma, P, S)$ — грамматика.
Базис. Если $a \in \Sigma$, то a — порождающий.
Индукция. Если $A \longrightarrow \alpha$ и цепочка α состоит только из порождающих (возможно, $\alpha = \varepsilon$), то A — порождающий.

Теорема А8.2.

Данный алгоритм находит в точности все порождающие грамматики $\mathfrak{G}$.

Упражнение А8.1.

Докажите теорему А8.2.

Алгоритм нахождения достижимых символов

Пусть $\mathfrak{G} = (V, \Sigma, P, S)$ — грамматика.
Базис. S — достижимый символ.
Индукция. Если $A \longrightarrow \alpha$ и A — достижимый символ, то все символы цепочки α также достижимы.

Алгоритм нахождения достижимых символов

Пусть $\mathfrak{G} = (V, \Sigma, P, S)$ — грамматика.
Базис. S — достижимый символ.
Индукция. Если $A \longrightarrow \alpha$ и A — достижимый символ, то все символы цепочки α также достижимы.

Теорема А8.3.

Данный алгоритм находит в точности все достижимые символы грамматики $\mathfrak{G}$.

Алгоритм нахождения достижимых символов

Вадим Пузаренко

Пусть $\mathfrak{G} = (V, \Sigma, P, S)$ — грамматика.
Базис. S — достижимый символ.
Индукция. Если $A \longrightarrow \alpha$ и A — достижимый символ, то все символы цепочки α также достижимы.

Теорема A8.3.

Данный алгоритм находит в точности все достижимые символы грамматики $\mathfrak{G}$.

Упражнение A8.2.

Докажите теорему A8.3.

Удаление ε-продукций

Определение А8.3.

Переменная A называется **ε-порождающей** в $\mathfrak{G}$, если $A \Rightarrow^*_{\mathfrak{G}} \varepsilon$.

Удаление ε-продукций

Определение А8.3.

Переменная A называется **ε-порождающей** в $\mathfrak{G}$, если $A \Rightarrow^*_{\mathfrak{G}} \varepsilon$.

Алгоритм.

Базис. Если $A \longrightarrow \varepsilon$ — продукция в $\mathfrak{G}$, то A — ε-порождающий.
Индукция. Если в $\mathfrak{G}$ есть продукция $B \longrightarrow C_1 C_2 \ldots C_k$, в которой C_i, $i = 1, \ldots, k$ — ε-порождающие, то и B также ε-порождающий.

Удаление ε-продукций

Определение A8.3.

Переменная A называется **ε-порождающей** в $\mathfrak{G}$, если $A \Rightarrow^*_{\mathfrak{G}} \varepsilon$.

Алгоритм.

Базис. Если $A \longrightarrow \varepsilon$ — продукция в $\mathfrak{G}$, то A — ε-порождающий.
Индукция. Если в $\mathfrak{G}$ есть продукция $B \longrightarrow C_1 C_2 \dots C_k$, в которой C_i, $i = 1, \dots, k$ — ε-порождающие, то и B также ε-порождающий.

Теорема A8.4.

В грамматике $\mathfrak{G}$ являются в точности ε-порождающими переменные, найденные вышеприведённым алгоритмом.

Удаление ε-продукций

Определение А8.3.

Переменная A называется ***ε-порождающей*** в $\mathfrak{G}$, если $A \Rightarrow^{*}_{\mathfrak{G}} \varepsilon$.

Алгоритм.

Базис. Если $A \longrightarrow \varepsilon$ — продукция в $\mathfrak{G}$, то A — ε-порождающий.
Индукция. Если в $\mathfrak{G}$ есть продукция $B \longrightarrow C_1 C_2 \ldots C_k$, в которой C_i, $i = 1, \ldots, k$ — ε-порождающие, то и B также ε-порождающий.

Теорема А8.4.

В грамматике $\mathfrak{G}$ являются в точности ε-порождающими переменные, найденные вышеприведённым алгоритмом.

Упражнение А8.3.

Докажите теорему А8.4.

Удаление ε-продукций

Конструкция.

1. Удаляем все продукции вида $A \longrightarrow \varepsilon$.
2. Если $A \longrightarrow \alpha\hat{\ }B\hat{\ }\gamma$ — продукция, в которой B — ε-порождающий, а $\mathrm{lh}(\alpha\hat{\ }\gamma) > 0$, то добавляем продукцию $A \longrightarrow \alpha\hat{\ }\gamma$. Повторяем данный пункт, пока это возможно.

Удаление ε-продукций

Конструкция.

1. Удаляем все продукции вида $A \longrightarrow \varepsilon$.
2. Если $A \longrightarrow \alpha\hat{\ }B\hat{\ }\gamma$ — продукция, в которой B — ε-порождающий, а $\mathrm{lh}(\alpha\hat{\ }\gamma) > 0$, то добавляем продукцию $A \longrightarrow \alpha\hat{\ }\gamma$. Повторяем данный пункт, пока это возможно.

Теорема A8.5.

Если грамматика $\mathfrak{G}_1$ построена по грамматике $\mathfrak{G}$ с помощью описанной выше конструкции удаления ε-продукций, то $L(\mathfrak{G}_1) = L(\mathfrak{G}) \setminus \{\varepsilon\}$.

Удаление ε-продукций

Конструкция.

1. Удаляем все продукции вида $A \longrightarrow \varepsilon$.
2. Если $A \longrightarrow \alpha\hat{\ }B\hat{\ }\gamma$ — продукция, в которой B — ε-порождающий, а $\mathrm{lh}(\alpha\hat{\ }\gamma) > 0$, то добавляем продукцию $A \longrightarrow \alpha\hat{\ }\gamma$. Повторяем данный пункт, пока это возможно.

Теорема A8.5.

Если грамматика $\mathfrak{G}_1$ построена по грамматике $\mathfrak{G}$ с помощью описанной выше конструкции удаления ε-продукций, то $L(\mathfrak{G}_1) = L(\mathfrak{G}) \setminus \{\varepsilon\}$.

Доказательство.

Необходимо доказать, что $\alpha \in L(\mathfrak{G}_1)$, если и только если $\alpha \in L(\mathfrak{G})$, для любого $\alpha \in \Sigma^* \setminus \{\varepsilon\}$. Докажем более общее утверждение ($\alpha \in (V \cup \Sigma)^*$): $A \Rightarrow^*_{\mathfrak{G}_1} \alpha \iff [A \Rightarrow^*_{\mathfrak{G}} \alpha \& (\alpha \neq \varepsilon)]$.

Удаление ε-продукций

Доказательство (продолжение).

($\Rightarrow$) Пусть $A \Rightarrow^*_{\mathfrak{G}_1} \alpha$; тогда $\alpha \neq \varepsilon$, поскольку $\mathfrak{G}_1$ не имеет ε-продукций. Покажем индукцией по длине порождения, что $A \Rightarrow^*_{\mathfrak{G}} \alpha$.

Базис. В $\mathfrak{G}_1$ имеется продукция $A \longrightarrow \alpha$; согласно конструкции, $A \longrightarrow \alpha'$ — продукция в $\mathfrak{G}$, где α получается из α' удалением ε-порождающих переменных. Следовательно, $A \Rightarrow_{\mathfrak{G}} \alpha' \Rightarrow^*_{\mathfrak{G}} \alpha$.

Индукция. Пусть в порождении $A \Rightarrow^*_{\mathfrak{G}_1} \alpha$ имеется $n > 1$ шагов. Тогда оно имеет вид $A \Rightarrow_{\mathfrak{G}_1} X_1X_2\ldots X_k \Rightarrow^*_{\mathfrak{G}_1} \alpha$. Цепочку α можно представить в виде $\alpha_1\hat{}\alpha_2\hat{}\ldots\hat{}\alpha_k$, где $X_i \Rightarrow^*_{\mathfrak{G}_1} \alpha_i$. По предположению индукции, $X_i \Rightarrow^*_{\mathfrak{G}} \alpha_i$. Согласно конструкции, в $\mathfrak{G}$ имеется продукция $A \longrightarrow Y_1Y_2\ldots Y_m$, где $X_1X_2\ldots X_k$ получена из $Y_1Y_2\ldots Y_m$ удалением ε-порождающих символов. Таким образом,

$$A \Rightarrow_{\mathfrak{G}} Y_1Y_2\ldots Y_m \Rightarrow^*_{\mathfrak{G}} X_1X_2\ldots X_k \Rightarrow^*_{\mathfrak{G}} \alpha_1\hat{}\alpha_2\hat{}\ldots\hat{}\alpha_k = \alpha.$$

Удаление ε-продукций

Доказательство (окончание).

$(\Leftarrow)$ Пусть $A \Rightarrow^*_{\mathfrak{G}} \alpha$ и $\alpha \neq \varepsilon$; как и прежде, будем доказывать индукцией по длине порождения.

Базис. $A \longrightarrow \alpha$ — продукция в $\mathfrak{G}$; так как $\alpha \neq \varepsilon$, $A \longrightarrow \alpha$ будет продукцией и в $\mathfrak{G}_1$. В частности, $A \Rightarrow^*_{\mathfrak{G}_1} \alpha$.

Индукция. Пусть в порождении $A \Rightarrow^*_{\mathfrak{G}} \alpha$ имеется $n > 1$ шагов. Тогда оно имеет вид $A \Rightarrow_{\mathfrak{G}} Y_1 Y_2 \ldots Y_m \Rightarrow^*_{\mathfrak{G}} \alpha$. Цепочку α можно представить в виде $\alpha_1 \hat{}\, \alpha_2 \hat{}\, \ldots \hat{}\, \alpha_m$, где $Y_i \Rightarrow^*_{\mathfrak{G}} \alpha_i$. Пусть цепочка $X_1 X_2 \ldots X_k$ получена из $Y_1 Y_2 \ldots Y_m$ удалением Y_j таких, что $\alpha_j = \varepsilon$. По предположению индукции, $Y_i \Rightarrow^*_{\mathfrak{G}_1} \alpha_i$. Согласно конструкции, $A \longrightarrow X_1 X_2 \ldots X_k$ — продукция в $\mathfrak{G}_1$. Таким образом, $A \Rightarrow_{\mathfrak{G}_1} X_1 X_2 \ldots X_k \Rightarrow^*_{\mathfrak{G}_1} \alpha_1 \hat{}\, \alpha_2 \hat{}\, \ldots \hat{}\, \alpha_m = \alpha$.

Для завершения доказательства осталось положить $A = S$. $\square$

Удаление цепных продукций

Определение A8.4.

Пара нетерминалов (A, B) назовём **цепной**, если $A \Rightarrow^* B$, причём в порождении используются только цепные продукции.

Удаление цепных продукций

Определение A8.4.

Пара нетерминалов (A, B) назовём **цепной**, если $A \Rightarrow^* B$, причём в порождении используются только цепные продукции.

Алгоритм.

1. (A, A) — цепная пара.
2. Если (A, B) — цепная пара и $B \longrightarrow C$ — цепная продукция, то (A, C) — цепная пара.

Удаление цепных продукций

Определение A8.4.

Пара нетерминалов (A, B) назовём **цепной**, если $A \Rightarrow^* B$, причём в порождении используются только цепные продукции.

Алгоритм.

1. (A, A) — цепная пара.
2. Если (A, B) — цепная пара и $B \longrightarrow C$ — цепная продукция, то (A, C) — цепная пара.

Теорема A8.6.

Вышеприведённый алгоритм находит в точности все цепные пары грамматики $\mathfrak{G}$.

Удаление цепных продукций

Определение А8.4.

Пара нетерминалов (A, B) назовём **цепной**, если $A \Rightarrow^* B$, причём в порождении используются только цепные продукции.

Алгоритм.

1. (A, A) — цепная пара.
2. Если (A, B) — цепная пара и $B \longrightarrow C$ — цепная продукция, то (A, C) — цепная пара.

Теорема А8.6.

Вышеприведённый алгоритм находит в точности все цепные пары грамматики $\mathfrak{G}$.

Упражнение А8.4.

Докажите теорему А8.6.

Удаление цепных продукций

Конструкция.

1. Найдём все цепные пары грамматики $\mathfrak{G}$.
2. Для каждой цепной пары (A, B) добавляем продукцию $A \longrightarrow \alpha$ в P_1, если $B \longrightarrow \alpha$ — нецепная продукция в P.

Удаление цепных продукций

Конструкция.

1. Найдём все цепные пары грамматики $\mathfrak{G}$.
2. Для каждой цепной пары (A, B) добавляем продукцию $A \longrightarrow \alpha$ в P_1, если $B \longrightarrow \alpha$ — нецепная продукция в P.

Замечание А8.1.

Заметим, что в случае, когда $A = B$, в P_1 помещаются все нецепные продукции из P.

Удаление цепных продукций

Конструкция.

1. Найдём все цепные пары грамматики $\mathfrak{G}$.
2. Для каждой цепной пары (A, B) добавляем продукцию $A \longrightarrow \alpha$ в P_1, если $B \longrightarrow \alpha$ — нецепная продукция в P.

Замечание А8.1.

Заметим, что в случае, когда $A = B$, в P_1 помещаются все нецепные продукции из P.

Теорема А8.7.

Если $\mathfrak{G}_1$ построена по $\mathfrak{G}$ согласно конструкции, описанной выше, то $L(\mathfrak{G}_1) = L(\mathfrak{G})$.

Удаление цепных продукций

Доказательство.

Докажем, что $\alpha \in L(\mathfrak{G}) \Leftrightarrow \alpha \in L(\mathfrak{G}_1)$.

Удаление цепных продукций

Доказательство.

Докажем, что $\alpha \in L(\mathfrak{G}) \Leftrightarrow \alpha \in L(\mathfrak{G}_1)$.
($\Leftarrow$) Пусть $S \Rightarrow^*_{\mathfrak{G}_1} \alpha$. Так как каждая продукция в $\mathfrak{G}_1$ эквивалентна последовательности из нескольких цепных продукций, за которыми следует одна нецепная продукция из $\mathfrak{G}$, имеем $S \Rightarrow^*_{\mathfrak{G}} \alpha$.

Удаление цепных продукций

Доказательство.

Докажем, что $\alpha \in L(\mathfrak{G}) \Leftrightarrow \alpha \in L(\mathfrak{G}_1)$.

($\Leftarrow$) Пусть $S \Rightarrow^*_{\mathfrak{G}_1} \alpha$. Так как каждая продукция в $\mathfrak{G}_1$ эквивалентна последовательности из нескольких цепных продукций, за которыми следует одна нецепная продукция из $\mathfrak{G}$, имеем $S \Rightarrow^*_{\mathfrak{G}} \alpha$.

($\Rightarrow$) Пусть $\alpha \in L(\mathfrak{G})$; по теореме A5.1, $S \underset{l}{\Rightarrow}^* \alpha$. Где бы в левом порождении ни использовалась цепная продукция, переменная её тела остаётся крайней слева. Тем самым, левое порождение в $\mathfrak{G}$ можно разбить на последовательность "шагов", в которых несколько цепных продукций сопровождаются нецепной. Заметим, что любая нецепная продукция, перед которой нет цепных, сама по себе образует такой "шаг". Но по построению грамматики $\mathfrak{G}_1$, каждый из этих шагов может быть выполнен одной её продукцией. Таким образом, $S \Rightarrow^*_{\mathfrak{G}_1} \alpha$. $\square$

Вспомогательные конструкции

Теорема А8.8.

Пусть $\mathfrak{G}$ — КС-грамматика, у которой $L(\mathfrak{G}) \setminus \{\varepsilon\} \neq \varnothing$. Тогда можно построить КС-грамматику $\mathfrak{G}_1$, в которой отсутствуют бесполезные символы, ε-продукции и цепные продукции, такую что $L(\mathfrak{G}_1) = L(\mathfrak{G}) \setminus \{\varepsilon\}$.

Вспомогательные конструкции

Теорема A8.8.

Пусть $\mathfrak{G}$ — КС-грамматика, у которой $L(\mathfrak{G}) \setminus \{\varepsilon\} \neq \varnothing$. Тогда можно построить КС-грамматику $\mathfrak{G}_1$, в которой отсутствуют бесполезные символы, ε-продукции и цепные продукции, такую что $L(\mathfrak{G}_1) = L(\mathfrak{G}) \setminus \{\varepsilon\}$.

Доказательство.

Сначала, по теореме A8.5, удалим ε-продукции; затем удалим цепные продукции, по теореме A8.7 (заметим, что в этом случае будут отсутствовать также и ε-продукции); в конечном итоге, применим конструкцию теоремы A8.1 (поскольку все продукции полученной грамматики содержатся во множестве продукций, полученной на предыдущем шаге, грамматика также не будет содержать ε- и цепных продукций). $\square$

Нормальная форма

Конструкция.

Опишем один шаг конструкции, в котором в грамматике происходят изменения только с одной продукцией. Пусть $A \longrightarrow X_1X_2\ldots X_k$, где либо $k > 2$, либо $k = 2$, но X_1 и X_2 не являются одновременно переменными. Для каждого слова X_1 и $X_2\ldots X_k$ вводим новую переменную Y_1 и Y_2, если оно не является нетерминалом. Возможны следующие случаи:

1. X_1 является нетерминалом; тогда рассматриваемую продукцию заменяем на следующий список: $A \longrightarrow X_1Y_2$, $Y_2 \longrightarrow X_2\ldots X_k$;
2. $X_2\ldots X_k = X_2$ является нетерминалом; тогда $A \longrightarrow Y_1X_2$, $Y_1 \longrightarrow X_1$;
3. X_1 является терминалом и $k > 2$; тогда $A \longrightarrow Y_1Y_2$, $Y_1 \longrightarrow X_1$, $Y_2 \longrightarrow X_2\ldots X_k$.

Конструкция завершится на конечном шаге (почему?)

Нормальная форма

Теорема А8.9.

Если $\mathfrak{G}$ — КС-грамматика, порождающая хотя бы одну непустую цепочку, то существует НФХ $\mathfrak{G}_1$ такая, что $L(\mathfrak{G}_1) = L(\mathfrak{G}) \setminus \{\varepsilon\}$.

Нормальная форма

Теорема A8.9.

Если $\mathfrak{G}$ — КС-грамматика, порождающая хотя бы одну непустую цепочку, то существует НФХ $\mathfrak{G}_1$ такая, что $L(\mathfrak{G}_1) = L(\mathfrak{G}) \setminus \{\varepsilon\}$.

Доказательство.

По теореме A8.8, можно построить КС-грамматику $\mathfrak{G}_2 = (V_2, \Sigma, P_2, S)$, свободную от бесполезных символов, ε-продукций и цепных продукций, для которой $L(\mathfrak{G}_2) = L(\mathfrak{G}) \setminus \{\varepsilon\}$.

Докажем, что $\alpha \in L(\mathfrak{G}_2) \Leftrightarrow \alpha \in L(\mathfrak{G}_1)$, где $\mathfrak{G}_1 = (V_1, \Sigma, P_1, S)$ строится по грамматике $\mathfrak{G}_2$ согласно конструкции.

$(\Rightarrow)$ Непосредственно вытекает из того, что $P_2(A, X_1X_2 \ldots X_k)$ влечёт $A \Rightarrow^*_{\mathfrak{G}_1} X_1X_2 \ldots X_k$.

$(\Leftarrow)$ Доказывать будем по длине вывода следующую импликацию: $A \Rightarrow^*_{l,\mathfrak{G}_1} \alpha \Longrightarrow A \Rightarrow^*_{l,\mathfrak{G}_2} \alpha$, для всех $A \in V_2$ и $\alpha \in \Sigma^*$.

Нормальная форма

Доказательство (продолжение).

Базис. Если $P_1(A,\alpha)$ и $A \in V_2$, то $\alpha \in \Sigma$ и $P_2(A,\alpha)$, поскольку $\mathfrak{G}_2$ не имеет цепных правил и ε-продукций.

Индукция. Пусть $A \Rightarrow^*_{l,\mathfrak{G}_1} \alpha$, и данное порождение имеет длину $n > 1$. Тогда $A \longrightarrow BC \Rightarrow^*_{l,\mathfrak{G}_1} \alpha$ и, следовательно, $B \Rightarrow^*_{l,\mathfrak{G}_1} \beta$, $C \Rightarrow^*_{l,\mathfrak{G}_1} \gamma$, причём $\alpha = \beta\hat{\ }\gamma$. Разберём несколько случаев.

- $B \notin V_2$; согласно конструкции, $P_1(B,\beta)$, $\beta \in \Sigma$ и $P_2(A, \beta X_2 \ldots X_k)$, причём эта продукция задаётся единственным образом. В частности, $A \Rightarrow^*_{l,\mathfrak{G}_2} \beta X_2 \ldots X_k$.
- $B \in V_2$; согласно конструкции, $P_2(A, BX_2 \ldots X_k)$, причём эта продукция задаётся единственным образом; по предположению индукции, $B \Rightarrow^*_{l,\mathfrak{G}_2} \beta$. В частности, $A \Rightarrow^*_{l,\mathfrak{G}_2} \beta X_2 \ldots X_k$.

Докажем теперь индукцией по k, что существует последовательность цепочек $\gamma_2, \ldots, \gamma_k$ такая, что $\gamma = \gamma_2\hat{\ }\ldots\hat{\ }\gamma_k$ и $X_i \Rightarrow^*_{l,\mathfrak{G}_2} \gamma_i$, $2 \leqslant i \leqslant k$.

Нормальная форма

Доказательство (окончание).

Базис. Возможны несколько случаев.

- $C \in V_2$; по индукционному предположению, $C = X_2 \Rightarrow^*_{l,\mathfrak{G}_2} \gamma$.
- $P_1(C,\gamma)$; тогда $\gamma \in \Sigma$ и, согласно конструкции, $P_2(A, B\gamma)$; в частности, $X_2 = \gamma \Rightarrow^*_{l,\mathfrak{G}_2} \gamma$.

Индукция. Пусть $P_1(C, C_1C_2)$. Разберём снова несколько случаев.

- $C_1 \in V_2$; тогда $X_2 = C_1 \Rightarrow^*_{l,\mathfrak{G}_2} \gamma_2$, а по предположению индукции, найдутся цепочки $\gamma_3, \ldots, \gamma_k$, удовлетворяющие условию.
- $C_1 \notin V_2$; тогда $P_2(C_1,\gamma_2)$, $\gamma_2 \in \Sigma$ и $X_2 = \gamma_2 \Rightarrow^*_{l,\mathfrak{G}_2} \gamma_2$; по предположению индукции, найдутся цепочки $\gamma_3, \ldots, \gamma_k$, удовлетворяющие условию.

Для завершения доказательства заметим, что $S \in V_2$ и, по доказанному, $\alpha \in L(\mathfrak{G}_2) \Leftrightarrow S \Rightarrow^*_{l,\mathfrak{G}_2} \alpha \Leftrightarrow S \Rightarrow^*_{l,\mathfrak{G}_1} \alpha \Leftrightarrow \alpha \in L(\mathfrak{G}_1)$. □

Спасибо за внимание.