ФОРМАЛЬНЫЕ ЯЗЫКИ И ГРАММАТИКИ

А. С. СААКЯН

НОРМАЛЬНЫЕ ФОРМЫ НЕКОТОРЫХ КЛАССОВ ГРАММАТИК ДЕРЕВЬЕВ

Грамматики деревьев (Δ-грамматики) в общем виде определены в работе [1]; там же указаны классы Δ-грамматик непосредственно составляющих и контекстно-свободных Δ-грамматик. Автоматные Δ-грамматики определены в [2]. Все перечисленные Δ-грамматики определялись аналогично соответствующим классам грамматик цепочек*, однако для линейных грамматик задача нахождения аналога в классе Δ-грамматик оставалась нерешенной.

В настоящей работе определяются понятия нормальных форм для контекстно-свободных и автоматных Δ-грамматик и строятся алгоритмы, приводящие эти Δ-грамматики к нормальным формам. Тем самым, с одной стороны, уточняется механизм порождения деревьев контекстно-свободными и автоматными Δ-грамматиками и, с другой стороны, становится возможным просто и естественно (и вновь по аналогии с грамматиками цепочек!) определить класс линейных Δ-грамматик, как промежуточный между классами контекстно-свободных и автоматных Δ-грамматик.

§ 1. Предварительные определения и обозначения

Объектами нашего изучения будут конечные ориентированные деревья (в дальнейшем для краткости они называются просто деревьями) с помеченными узлами.

Деревья t_1 и t_2 назовем изоморфными, если между множествами узлов этих деревьев можно установить одно-однозначное соответствие так, что:

а) соответственные узлы деревьев t_1 и t_2 помечены одними и теми же символами;

б) в дереве t_1 из узла α идет дуга в узел β тогда и только тогда, когда в дереве t_2 из образа узла α идет дуга в образ узла β.

В дальнейшем мы не различаем изоморфные между собой деревья.

---

* Грамматикам цепочек посвящена, например, монография: *Гладкий А. В.* Формальные грамматики и языки, М., «Наука», 1973.

Через T_V будем обозначать множество всех деревьев, узлы которых помечены элементами некоторого конечного множества V. Всякое подмножество множества T_V назовем языком деревьев (Δ-языком) над словарем V.

Будем говорить, что дерево t_2 является α-поддеревом дерева t_1, если t_2 является связным подграфом t_1 и корнем дерева t_2 является узел α дерева t_1. Скажем, что дерево t_2 является полным α-поддеревом дерева t_1, если t_2 изоморфно дереву, полученному из t_1 удалением всех отличных от α узлов, в которые не идут пути из α, и инцидентных им дуг.

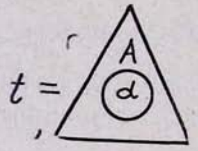
Скажем, что дерево t_2 является квазиполным α-поддеревом дерева t_1, если t_2 есть α-поддерево дерева t_1 и всякое полное β-поддерево, $\beta \neq \alpha$, дерева t_2 является полным β-поддеревом дерева t_1.

Будем говорить, что дерево t_2 является поддеревом (полным поддеревом, квазиполным поддеревом) дерева t_1, если оно является α-поддеревом (соответственно полным α-поддеревом, квазиполным α-поддеревом) дерева t_1 для некоторого узла α из t_1.

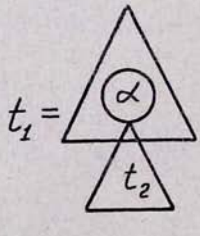
Узлы дерева, не являющиеся началами никаких дуг, назовем висячими. Минимальным деревом назовем дерево, все узлы которого, кроме, быть может, корня, висячие; частным случаем минимального дерева является дерево, состоящее из одного лишь корня—единичное дерево.

Высотой узла дерева назовем длину пути из корня в данный узел. Наибольшее значение высоты узла дерева назовем высотой дерева. Высота дерева t будет обозначаться через $h(t)$.

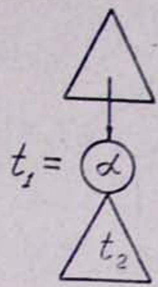
Условимся о следующих изображениях деревьев. Если в дереве t выделен узел α с пометкой A, то это мы изобразим так, как на схеме 1;



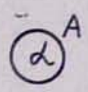
если t_2 — квазиполное α-поддерево дерева t_1, то это будет изображаться так, как на схеме 2;



наконец, если t_2 — полное α-поддерево дерева t_1, то это будет изображаться так, как на схеме 3.



В дальнейшем, когда это не вызывает недоразумений, мы будем отождествлять узел с его меткой; более того, если t-единичное дерево с единственным узлом α, помеченным символом A, то вместо изображения, указанного на схеме 4,



мы будем писать один только символ A.

Сейчас мы введем две операции над деревьями: композицию и квазикомпозицию.

Операция композиции определяется следующим образом. Пусть дано дерево t_0 и n деревьев $t_1, t_2, \ldots, t_n$; пусть, кроме того, в дереве t_0 выделено n попарно различных узлов $\alpha_1, \alpha_2, \ldots, \alpha_n$. Тогда результатом композиции дерева t_0 с выделенными узлами $\alpha_1, \alpha_2, \ldots, \alpha_n$ и деревьев $t_1\ t_2, \ldots, t_n$ будет, по определению, дерево T, полученное из t_0 „наклеиванием" корней деревьев $t_1, \ldots, t_n$ на его узлы $\alpha_1, \ldots, \alpha_n$ соответственно; при этом, для каждого $i = 1, 2, \ldots, n$ пометка в узле α_i результирующего дерева T совпадает с пометкой в корне дерева t_i (а не с пометкой в узле α_i дерева t_0!).

Композицию дерева t_0 с выделенными узлами $\alpha_1, \alpha_2, \ldots, \alpha_n$ и деревьев $t_1, t_2, \ldots, t_n$ будем обозначать через

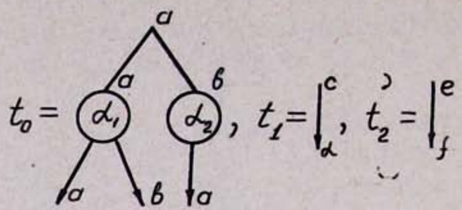
$$K(t_0; \alpha_1, \alpha_2, \ldots, \alpha_n \mid t_1, t_2, \ldots, t_n).$$

Теперь определим операцию квазикомпозиции. Пусть дано дерево t_0 и n деревьев $t_1, t_2, \ldots, t_n$: пусть, кроме того, в дереве t_0 выделено n узлов $\alpha_1, \alpha_2, \ldots, \alpha_n$ не обязательно попарно различных. Тогда результатом квазикомпозиции дерева t_0 с выделенными узлами $\alpha_1, \alpha_2, \ldots, \alpha_n$ и деревьев $t_1, t_2, \ldots, t_n$ будет, по определению, дерево T', полученное из t_0 „наклеиванием" корней деревьев $t_1, t_2, \ldots, t_n$ на его узлы $\alpha_1, \alpha_2, \ldots, \alpha_n$ соответственно; при этом для каждого $i = 1, 2, \ldots, n$ пометка в узле α_i результирующего дерева T' совпадает с пометкой узла α_i в дереве t_0 (а не с пометкой корня дерева t_i!).

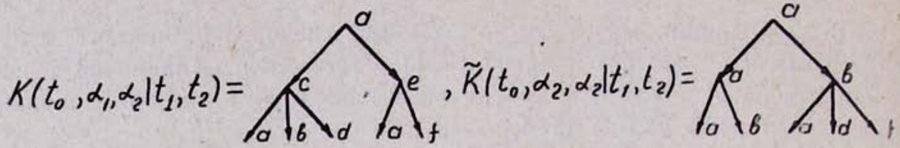
Квазикомпозицию дерева t_0 с выделенными узлами $\alpha_1, \alpha_2, \ldots, \alpha_n$ и деревьев $t_1, t_2, \ldots, t_n$ будем обозначать через

$$\bar{K}(t_0; \alpha_1, \alpha_2, \ldots, \alpha_n \mid t_1, t_2, \ldots, t_n).$$

Пример. Пусть заданы деревья, изображенные на схеме 5.



Операции композиции и квазикомпозиции иллюстрирует схема 6.



§ 2. Грамматики деревьев

(Порождающая) грамматика деревьев (Δ-грамматика) есть упорядоченная четверка $\Gamma = \langle V, W, I, R \rangle$ где: 1) V; 2) W — непересекающиеся непустые конечные множества; 3) I — некоторый элемент W; 4) R — конечное множество упорядоченных троек $\langle t_1, t_2, f \rangle$, где t_1 и t_2 — деревья над множеством $V \cup W$, а f — однозначное отображение множества узлов дерева t_1 во множество узлов дерева t_2. Множества V и W называются, соответственно, основным и вспомогательным словарями Δ-грамматики Γ, а их элементы — соответственно, основными и вспомогательными символами; I называется начальным символом Γ, R-схемой Γ и элементы R-правилами Γ.

Вместо $\langle t_1, t_2, f \rangle$ будет использоваться, для большей наглядности, запись $t_1 \to t_2 | f$, а деревья t_1 и t_2 будут называться соответственно левой и правой частями этого правила.

Для левой и правой частей правила $p = t_1 \to t_2 | f$ мы будем пользоваться стандартными обозначениями: $l(p) = t_1$ и $r(p) = t_2$. Таким образом, всякое правило вывода p запишется в виде $l(p) \to r(p) | f$.

Узлы дерева t_2, имеющие при отображении f прообразы в t_1, назовем крючками.

Будем говорить, что дерево T' получается из дерева T применением правила $t_1 \to t_2 | f$, если T и T' представимы соответственно в виде

$$T = K(T_0;\ \alpha_0 | K(t_1;\ \alpha_1, \alpha_2, \ldots, \alpha_n | T_1, T_2, \ldots, T_n)) \tag{1}$$

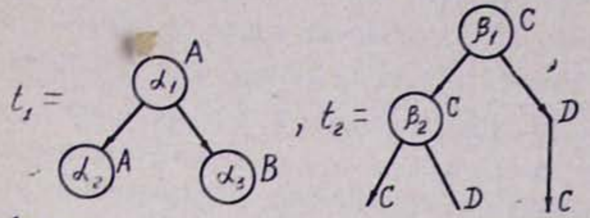
и

$$T' = K(T_0;\ \alpha_0 | \tilde{K}(t_2;\ f(\alpha_1), f(\alpha_2), \ldots, f(\alpha_n) | T_1, T_2, \ldots, T_n)), \tag{2}$$

где α_0 — висячий узел дерева T_0, а $\alpha_1, \alpha_2, \ldots, \alpha_n$ — бесповторный пересчет всех узлов дерева t_1*.

* Подчеркнем, что узлы $\alpha_1, \alpha_2, \ldots, \alpha_n$ попарно различны, а $f(\alpha_1), f(\alpha_2), \ldots, f(\alpha_n)$ не обязательно попарно различны.

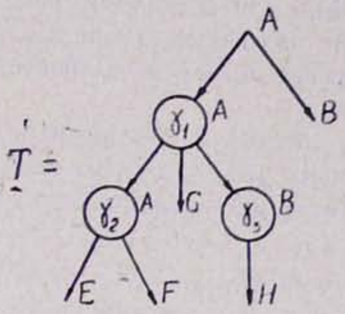
Пример. Пусть имеется правило $t_1 \to t_2 | f$, где t_1 и t_2 имеют вид, изображенный на схеме 7,



а f задано таблицей

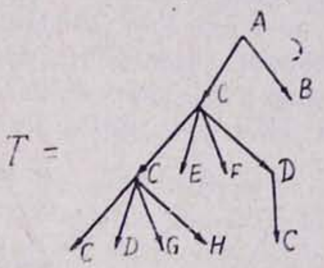
$$f(\alpha_1) = f(\alpha_2) = \beta_2,$$
$$f(\alpha_2) = \beta_1.$$

Применим указанное правило к дереву T, заданному схемой 8.



В дереве T имеется два поддерева t_1.

Применим указанное правило к поддереву t_1, имеющему узлы $\gamma_1, \gamma_2, \gamma_3$; в результате получится дерево T', изображенное на схеме 9.



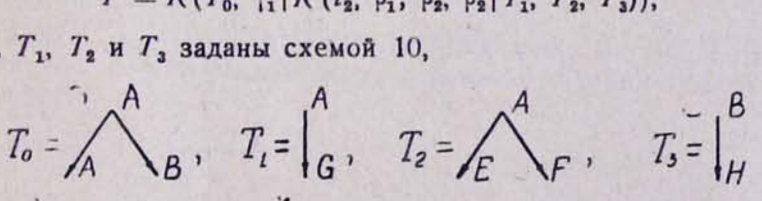
Представления (1) и (2) будут иметь здесь вид

$$T = K(T_0;\ \gamma_1 | K(t_1;\ \alpha_1,\ \alpha_2,\ \alpha_3 | T_1,\ T_2,\ T_3))$$

и

$$T' = K(T_0;\ \gamma_1 | \widetilde{K}(t_2;\ \beta_1,\ \beta_2,\ \beta_2 | T_1,\ T_2,\ T_3)),$$

где T_0, T_1, T_2 и T_3 заданы схемой 10,



Пусть задана Δ-грамматика Γ. Если дерево T' получается из дерева T применением какого-либо правила Γ, будем говорить, что T' непосредственно выводимо из T в Γ, и писать $T \overset{*}{\underset{\Gamma}{\Rightarrow}} T'$. Будем говорить, что дерево T' выводимо из дерева T и писать $T \underset{\Gamma}{\Rightarrow} T'$, если либо $T = T'$, либо существует конечная последовательность деревьев $T_1, T_2, \ldots, T_m$ $(m > 0)$ такая, что $T \overset{*}{\underset{\Gamma}{\Rightarrow}} T_1 \overset{*}{\underset{\Gamma}{\Rightarrow}} T_2 \overset{*}{\underset{\Gamma}{\Rightarrow}} \ldots \overset{*}{\underset{\Gamma}{\Rightarrow}} T_m = T'$.

Число m назовем длиной вывода T' из T в Δ-грамматике Γ.

Языком деревьев (Δ-языком), порождаемым Δ-грамматикой $\Gamma = \langle V, W, I, R \rangle$, назовем множество

$$L(\Gamma) = \{t \mid I \underset{\Gamma}{\Rightarrow} t, \ t \in T_V\}.$$

Δ-грамматики Γ и Γ' назовем эквивалентными, если $L(\Gamma) = L(\Gamma')$.

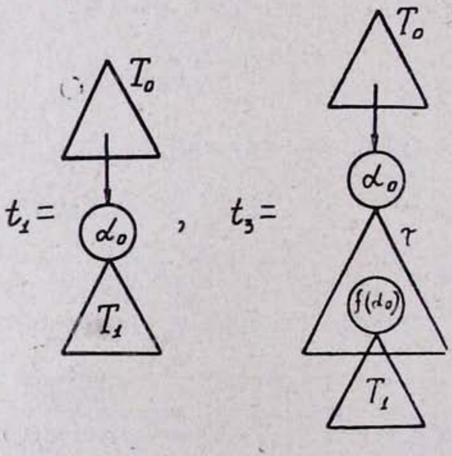
Порождающие Δ-грамматики являются исчислениями весьма общего вида, в то время как для возможных приложений представляют, по-видимому, интерес их различные специальные классы. Специализировать классы Δ-грамматик можно, в частности, путем наложения ограничений на правила.

Δ-грамматику $\Gamma = \langle V, W, I, R \rangle$ назовем Δ-грамматикой непосредственно составляющих (НС — Δ-грамматикой), если для каждого ее правила $t_1 \to t_2 | f$ имеет место следующее представление деревьев t_1 и t_2:

$$t_1 = K(T_0;\ \alpha_0 | T_1),$$

$$t_2 = K(T_0;\ \alpha_0 | \widetilde{K}(\tau;\ f(\alpha_0) | T_1)),$$

где: 1) α_0 — висячий узел дерева T_0, помеченный вспомогательным символом; 2) $f(\alpha_0)$ — узел дерева τ; 3) на всех узлах дерева t_1, отличных от α, отображение f является тождественным. Правила, удовлетворяющие этим условиям, назовем НС-правилами. Наглядно деревья t_1 и t_2, участвующие в НС-правиле, можно представить в виде, изображенном на схеме 11.



В НС-правиле $t_1 \to t_2 | f$ деревья T_0 и T_1 естественно называть соответственно верхним и нижним контекстами. При применении НС-правила фактически заменяется только один узел, но возможность замены зависит от наличия нужного контекста.

Сделаем следующий шаг на пути специализации классов Δ-грамматик. Потребуем, чтобы левые части правил НС-Δ-грамматик состояли только из единичных деревьев, т. е. чтобы контекст был пуст. Сформулируем точное определение.

Δ-грамматику $\Gamma = \langle V, W, I, R \rangle$ назовем контекстно-свободной (КС—Δ-грамматикой), если для каждого правила $p \in R$, $l(p)$ представляет собой единичное дерево, помеченное вспомогательным символом. Правила, удовлетворяющие этому условию, назовем КС-правилами. КС-правило p мы часто, в соответствии с договоренностью, принятой в § 1, будем записывать в виде $A \to r(p) | f$, где A — вспомогательный символ, которым помечено единичное дерево $l(p)$.

Высотой КС-правила p назовем величину $h(p) = h(r(p))$; высотой КС—Δ-грамматики $\Gamma = \langle V, W, I, R \rangle$ назовем величину $h(\Gamma) = \max_{p \in R} h(p)$.

КС-правило p назовем линейным (Л-правилом), если узел $f(l(p))$ совпадает с корнем дерева $r(p)$. Δ-грамматику назовем линейной (Л—Δ-грамматикой), если все ее правила линейные.

Договоримся опускать знак функции в Л-правилах, так как она в любом Л-правиле производит одно и то же отображение. Таким образом, Л-правило p запишется в виде $A \to r(p)$, где A — метка $l(p)$.

Л-правило p назовем автоматным (А-правилом), если корень дерева $r(p)$ помечен основным символом, а Δ-грамматику, содержащую только А-правила, назовем автоматной Δ-грамматикой (А-Δ-грамматикой). Δ-языки, порождаемые НС-, КС-, Л-, и А-Δ-грамматиками, назовем соответственно НС-, КС-, Л- и А-Δ-языками.

§ 3. Нормальные формы КС- и Л- Δ-грамматик

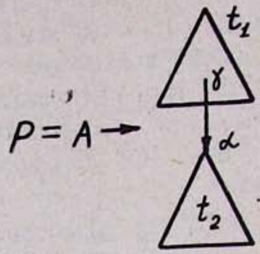
Будем говорить, что КС-Δ-грамматика $\Gamma = \langle V, W, I, R \rangle$ находится в нормальной форме, если каждое правило из R имеет один из двух видов, изображенных на схеме 12:

$$A \to \begin{array}{c} \alpha \\ \downarrow \\ \beta \end{array} | f, \quad A \to \alpha$$

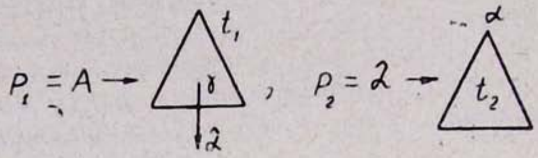
где $A \in W$ и $\alpha, \beta \in V \cup W$.

Теорема 3.1. Для каждой КС- Δ-грамматики $\Gamma = \langle V, W, I, R \rangle$ можно эффективно построить эквивалентную КС-Δ-грамматику в нормальной форме.

Доказательство. 1) Пусть p (см. схему 13)



— некоторое правило Г, где α, $\gamma \in V \cup W$ и α-крючок. Построим два новых КС-правила — p_1 и p_2 (см. схему 14),

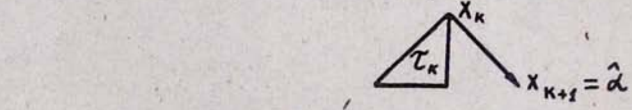


где $\hat{\alpha} \bar{\in} V \cup W$, $\hat{\alpha}$ — крючок правила p_1 и α — крючок правила p_2. Положим

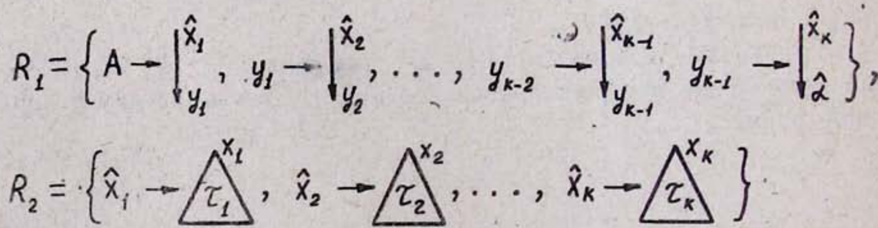
$$\Gamma^1 = \langle V,\ W \cup \{\hat{\alpha}\},\ I,\ (R \setminus \{p\}) \cup \{p_1,\ p_2\} \rangle .$$

Очевидно, что Γ^1 есть КС- Δ-грамматика, эквивалентная Г.

Рассмотрим правило p_1 из Γ^1. В дереве $r(p_1)$ выделим путь, идущий из корня в висячий узел $\hat{\alpha}$; тогда для некоторых $x_1, x_2, \ldots, x_k \in V \cup W$ и некоторых деревьев $\tau_1, \tau_2, \ldots, \tau_n$ правило p_1 можно изобразить так, как это сделано на схеме 15.



Построим два множества КС-правил R_1 и R_2 (см. схему 16),

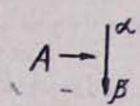


где $\hat{x}_1, \hat{x}_2, \ldots, \hat{x}_k, y_1, y_2, \ldots, y_{k-1} \in V \cup W$; $y_1, y_2, \ldots, y_{k-1}, \hat{\alpha}$ — крючки правил из R_1 и $x_1, x_2, \ldots, x_k$ — крючки правил из R_2. Ясно, что КС-Δ-грамматика

$$\Gamma_2 = \langle V,\ W \cup \{\hat{\alpha}, \hat{x}_1, \hat{x}_2, \ldots, \hat{x}_k, y_1, y_2, \ldots, y_{k-1}\},$$
$$I,\ (R \setminus \{p\}) \cup \{p_2\} \cup R_1 \cup R_2 \rangle$$

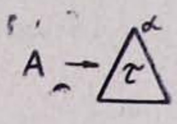
эквивалентна Γ^1.

2) Применим процедуру пункта 1 к каждому правилу из R. В результате не более чем $\mu(R)$* шагов будет построена КС- Δ-грамматика $\Gamma_0 = \langle V, W, I, R_0 \rangle$, эквивалентная Γ и такая, что каждое правило из R_0 представимо либо в виде, изображенном на схеме 17



(I)

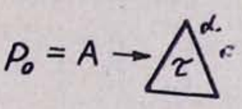
где $A \in W$, α, $\beta \in V \cup W$ и β-крючок, либо в виде, изображенном на на схеме 18,



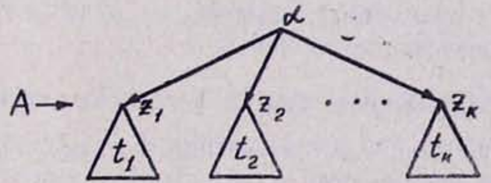
(II)

где $A \in W$, $\alpha \in V \cup W$ и α-крючок.

3) Пусть p_0 (см. схему 19)



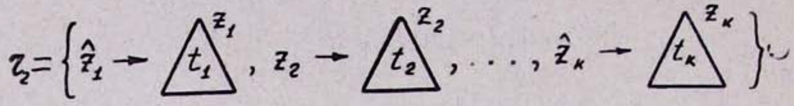
— какое-нибудь правило вида (II) из R_0. Для некоторых $z_1, z_2, \ldots, z_k \in V \cup W$ и некоторых деревьев $t_1, t_2, \ldots, t_k$ правило p_0 можно изобразить, как на схеме 20.



* Мы пользуемся принятым обозначением $\mu(M)$ для мощности множества M.

Построим два множества КС-правил r_1 и r_2 (см. схему 21),

$$r_1=\left\{A\to\begin{array}{l}\alpha_1\\ \big\downarrow\\ \hat{z}_1\end{array},\ \alpha_1\to\begin{array}{l}\alpha_2\\ \big\downarrow\\ \hat{z}_2\end{array},\ \ldots,\ \alpha_{k-2}\to\begin{array}{l}\alpha_{k-1}\\ \big\downarrow\\ \hat{z}_{k-1}\end{array},\ \alpha_{k-1}\to\begin{array}{l}\alpha\\ \big\downarrow\\ \hat{z}_k\end{array}\right\},$$



где $\alpha_1, \alpha_2, \ldots, \alpha_{k-1}, \hat{z}_1, \hat{z}_2, \ldots, \hat{z}_k \in V \cup W$, а крючками правил множеств r_1 и r_2 являются корни правых частей правил. Ясно, что КС-Δ-грамматика

$$\Gamma_0' = \langle V,\ W \cup \{\alpha_1, \ldots, \alpha_{k-1}, \hat{x}_1, \ldots, \hat{x}_k\},\ I,\ (R \setminus \{p_0\} \cup r_1 \cup r_2 \rangle$$

эквивалентна Γ_0.

Применив описанное преобразование к каждому правилу вида (II) из R_0, получим КС- Δ-грамматику Γ_0'', эквивалентную Γ_0 и такую, что $h(\Gamma_0'') < \hat{h}(\Gamma_0)$,

4) $\mu(h(\Gamma_0))$-кратное применение процедуры пункта 3 приведет к построению КС- Δ-грамматики в нормальной форме $\hat{\Gamma}$, эквивалентной Γ_0.

Теорема доказана.

Будем говорить, что Л- Δ-грамматика $\Gamma = \langle V, W, I, R \rangle$ находится в нормальной форме, если каждое правило из R имеет один из двух видов, изображенных на схеме 22

$$A\to\begin{array}{l}\alpha\\ \big\downarrow\\ \beta\end{array},\quad A\to\alpha$$

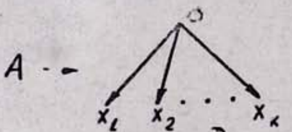
где $A \in W$ и $\alpha, \beta \in V \cup W$.

Т е о р е м а 3.2. Для каждой Л- Δ-грамматики Γ можно эффективно построить эквивалентную Л- Δ-грамматику в нормальной форме.

Д о к а з а т е л ь с т в о этой теоремы по существу заключено в пунктах 3 и 4 доказательства теоремы 3.1.

§ 4. Нормальная форма А- Δ-грамматик

Будем говорить, что А- Δ-грамматика $\Gamma = \langle V, W, I, R \rangle$ находится в нормальной форме, если $h(\Gamma) \leqslant 1$. Другими словами, каждое правило А- Δ-грамматики в нормальной форме имеет вид, изображенный на схеме 23,



для некоторых $A \in W$, $a \in V$ и $x_1, x_2, \ldots, x_k \in V \cup W$.

Теорема 4.1. Для всякой А- Δ-грамматики $\Gamma = \langle V, W, I, R \rangle$ можно эффективно построить эквивалентную А- Δ-грамматику Γ_0 в в нормальной форме.

Доказательству этой теоремы мы предпошлем несколько новых обозначений и определений.

Для произвольного дерева t пусть $g(t)$ — множество всех узлов t, высоты которых равны $h(t)-1$. Для А- Δ-грамматики $\Gamma = \langle V, W, I, R \rangle$ положим:

$$M(\Gamma) = \max_{p \in R,\ h(p) = h(\Gamma)} \mu(g(r(p))),$$

$$N(\Gamma) = \mu(\{p \mid p \in R,\ h(p) = h(\Gamma),\ \mu(g(r(p))) = M(\Gamma)\}).$$

Сложностью А- Δ-грамматики Γ назовем упорядоченную тройку $S(\Gamma) = \langle h(\Gamma), M(\Gamma), N(\Gamma) \rangle$.

По определению полагаем $S(\Gamma_1) < S(\Gamma_2)$, если тройка $\langle h(\Gamma_1), M(\Gamma_1), N(\Gamma_1) \rangle$ лексикографически предшествует тройке $\langle h(\Gamma_2), M(\Gamma_2), N(\Gamma_2) \rangle$.

Для произвольных дерева t и множества W обозначим

$$g_W(t) = \{\alpha \mid \alpha \in g(t),\ \text{узел } \alpha \text{ помечен элементом из } W\}.$$

Для А- Δ-грамматики $\Gamma = \langle V, W, I, R \rangle$ положим:

$$U_\Gamma = \bigcup_{p \in R,\ h(p) = h(\Gamma)} g_W(r(p)),$$

$$W_g^\Gamma = \{A \mid A \in W_g^\Gamma,\ A \text{ — метка какого-либо узла из } U_\Gamma\},$$

$$R_g^\Gamma = \{p \mid p \in R,\ h(p) = h(\Gamma),\ l(p) \in W_g^\Gamma,\ g_W(r(p)) \neq \varnothing\}.$$

g — сложностью А- Δ-грамматики Γ назовем упорядоченную тройку

$$S_g(\Gamma) = \langle h_g(\Gamma), M_g(\Gamma), N_g(\Gamma) \rangle,$$

где

$$h_g(\Gamma) = \begin{cases} h(\Gamma), & \text{если } R_g(\Gamma) \neq \varnothing, \\ 0, & \text{если } R_g(\Gamma) = \varnothing, \end{cases}$$

$$M_g(\Gamma) = \max_{p \in R_g^\Gamma} \mu(g_W(r(p))),$$

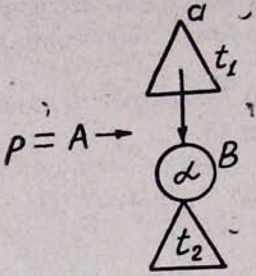
$$N_g(\Gamma) = \mu(\{p \mid p \in R_g^\Gamma,\ \mu(g_W(r(p))) = M_g(\Gamma)\}).$$

Отношение $<$ для g-сложностей определяется так же, как и для сложностей — лексикографически.

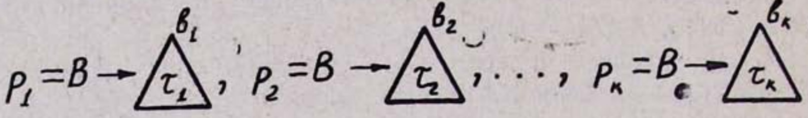
Легко понять, что $R_g^\Gamma = \varnothing$ тогда и только тогда, когда $S_g(\Gamma) = \langle 0, 0, 0 \rangle$.

Лемма 4.1. По всякой А- Δ-грамматике $\Gamma = \langle V, W, I, R \rangle$, для которой $S_g(\Gamma) \neq \langle 0, 0, 0 \rangle$, можно эффективно построить эквивалентную А- Δ-грамматику Γ' такую, что $S_g(\Gamma') < S_g(\Gamma)$.

Доказательство. Так как $S_g(\Gamma) \neq \langle 0, 0, 0 \rangle$, то $R_g^\Gamma \neq \varnothing$; пусть p (см. схему 24) — некоторый элемент R_g^Γ,



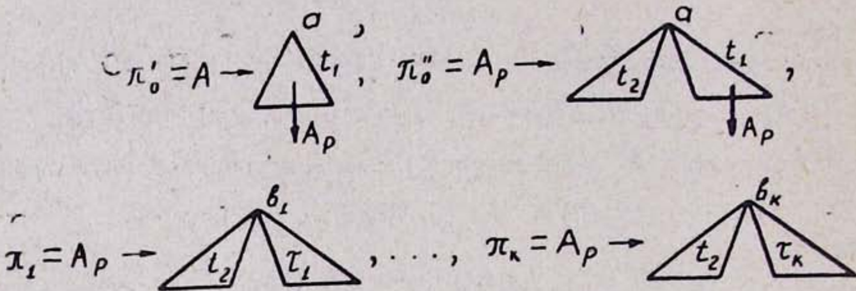
где $a \in V$, A, $B \in W_g^{\Gamma}$ и $\alpha \in g_W(r(p))$ (т. е. $h(t_2) = 1$). Пусть, далее, $p_1, p_2, \ldots, p_k$, $k > 0$* (см. схему 25),



— все те правила из R, в левых частях которых стоит символ B и $b_1, b_2, \ldots, b_k \in V$.

Нам следует отдельно рассмотреть два случая — $A \neq B$ и $A = B$.

1) $A \neq B$. Исключим из R правило p и вместо него добавим $k+1$ новых правил $\pi_0, \pi_1, \pi_2, \ldots, \pi_k$ (см. схему 26),



где $A_p \overline{\in} V \cup W$. Очевидно, что А-Δ-грамматика

$$\Gamma' = \langle V,\ W \cup \{A_p\},\ I,\ (R \setminus \{p\}) \cup \{\pi_0, \pi_1, \pi_2, \ldots, \pi_k\} \rangle$$

эквивалентна Г.

Покажем, что $S_g(\Gamma') < S_g(\Gamma)$.

Возможны два случая $h_g(\Gamma') < h_g(\Gamma)$ и $h_g(\Gamma') = h_g(\Gamma)$.

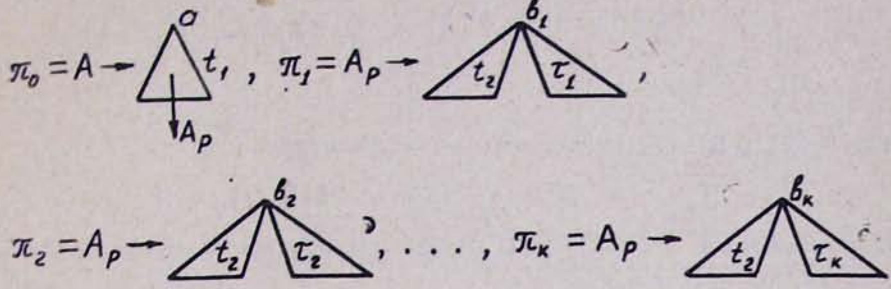
Если $h_g(\Gamma') < h_g(\Gamma)$, то сразу $S_g(\Gamma) < S_g(\Gamma)$. Пусть $h_g(\Gamma') = h_g(\Gamma)$; тогда либо (при $h(t_1) < h(\Gamma)$ или $g_W(t_1) = \varnothing$) $R_g^{\Gamma'} = R_g^{\Gamma} \setminus \{p\}$, либо при $h(t_1) = h(\Gamma)$ и $g_W(t_1) \neq \varnothing$) $R_g^{\Gamma'} = (R_g^{\Gamma} \setminus \{p\} \cup \{\pi_0\}$** и, следова-

* Если $k = 0$, то из R можно удалить правило p, не изменив при этом порождающей способности Г.

** Ни одно из правил $\pi_1, \pi_2, \ldots, \pi_k$ в множество $R_g^{\Gamma'}$ не попадет, так как $A_p \overline{\in} W_g^{\Gamma'}$.

тельно, либо $M_g(\Gamma') < M_g(\Gamma)$, либо, если $M_g(\Gamma') = M_g(\Gamma)$, то, с необходимостью, $N_g(\Gamma') < N_g(\Gamma)$. Таким образом, в любом случае $S_g(\Gamma') < S_g(\Gamma)$.

2) $A = B$. Исключим из R правило p и вместо него добавим $k+2$ новых правил: π_0' π_0'', $\pi_1, \ldots, \pi_k$ (см. схему 27),



где $A_p \notin V \cup W$. Понятно, что А- Δ-грамматика

$$\Gamma' = \langle V,\ W \cup \{A_p\},\ I,\ (R \setminus \{p\}) \cup \{\pi_0',\ \pi_0'',\ \pi_1, \ldots, \pi_k\} \rangle$$

эквивалентна Г.

Покажем, что и в этом случае $S_g(\Gamma') < S_g(\Gamma)$.

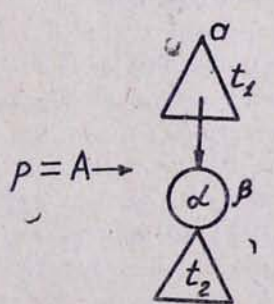
Если $h_g(\Gamma') < h_g(\Gamma)$, то сразу и $S_g(\Gamma') < S_g(\Gamma)$. Пусть $h_g(\Gamma') = h_g(\Gamma)$; тогда, аналогично пункту 1), либо $R_g^{\Gamma'} = R_g^{\Gamma} \setminus \{p\}$, либо $R_g^{\Gamma'} = (R_g^{\Gamma} \setminus \{p\}) \cup \{\pi_0''\}$, и, следовательно, либо $M_g(\Gamma') < M_g(\Gamma)$, либо, если $M_g(\Gamma') = M_g(\Gamma)$, то, с необходимостью, $N_g(\Gamma') < N_g(\Gamma)$, т. е. в любом случае $S_g(\Gamma') < S_g(\Gamma)$. Лемма доказана.

Следствие 4.1. По всякой А- Δ-грамматике Г можно эффективно построить эквивалентную А- Δ-грамматику Г′, для которой $S_g(\Gamma') = \langle 0, 0, 0 \rangle$.

Замечание. В только что доказанной лемме и ее следствии сложность $S(\Gamma')$ вообще говоря, может возрасти по сравнению с $S(\Gamma)$ однако всегда $h(\Gamma') \leqslant h(\Gamma)$.

Лемма 4.2. По всякой А- Δ-грамматике $\Gamma = \langle V, W, I, R \rangle$, для которой $h(\Gamma) > 1$ и $S_g(\Gamma) = \langle 0, 0, 0 \rangle$, можно эффективно построить эквивалентную А- Δ-грамматику Γ_1 такую, что $S(\Gamma_1) < S(\Gamma)$.

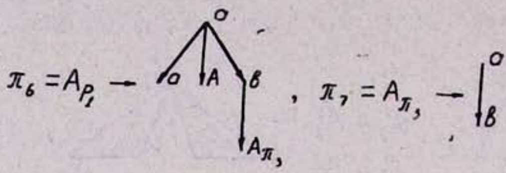
Доказательство. Пусть $p \in R$, $h(p) = h(\Gamma)$ $\mu(g(r(p))) = M(\Gamma)$



(правило p изображено на схеме 28, где $a \in V$, $A \in W$, $\beta \in V \cup W$ $\alpha \in g(r(p))$.

Рассмотрим отдельно два случая: $\beta \in V$ и $\beta \in W$.

1) $\beta = b \in V$. Исключим из R правило p и вместо него добавим два новых правила: π_1 и π_2 (см. схему 29),

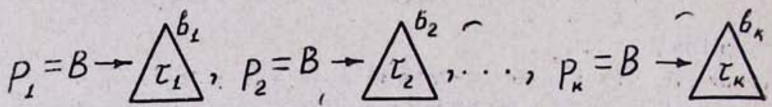


где $A_p \overline{\in} V \cup W$. Очевидно, что А- Δ-грамматика

$$\Gamma_1 = \langle V,\ W \cup \{A_p\},\ I,\ (R \setminus \{p\}) \cup \{\pi_1, \pi_2\} \rangle$$

эквивалентна Γ и $S(\Gamma_1) < S(\Gamma)$; заметим при этом, что если $h(\Gamma_1) = h(\Gamma)$, то $S_g(\Gamma_1) = \langle 0, 0, 0 \rangle$.

2) $\beta = B \in W$. Пусть $p_1, p_2, \ldots, p_k$ (см. схему 30) — все те правила



из R, в левых частях которых символ B; при этом, так как $S_g(\Gamma) = \langle 0, 0, 0 \rangle$, то $R_g^\Gamma = \varnothing$ и, следовательно, либо $h(p_i) < h(\Gamma)$, либо $g_W(r(p_i)) = \varnothing$, для $i = 1, 2, \ldots, k$. Если для некоторого i, $g_W(r(p_i)) = \varnothing$, но $h(p_i) = h(\Gamma)$, то применением к p_i процедуры пункта 1 можно легко добиться укорочения этого правила; поэтому с самого начала положим $h(p_i) < h(\Gamma)$.

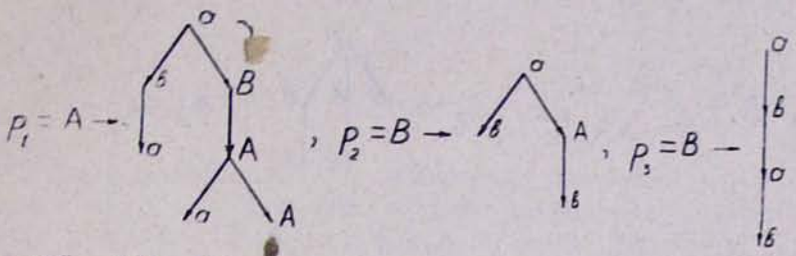
Очевидно, что $A \neq B$, иначе $R_g^\Gamma \neq \varnothing$. Возьмем $\Gamma_1 = \Gamma'$, где Γ' — А- Δ-грамматика, построенная в пункте 1 леммы 4.1. Ясно, что Γ_1 эквивалентна Γ и $S(\Gamma_1) < S(\Gamma)$; заметим при этом, что если $h(\Gamma_1) = h(\Gamma)$, то $S_g(\Gamma_1) = \langle 0, 0, 0 \rangle$. Лемма доказана.

Следствие 4.2. По всякой А- Δ-грамматике Γ, для которой $h(\Gamma) > 1$ и $S_g(\Gamma) = \langle 0, 0\ 0, \rangle$, можно эффективно построить эквивалентную А- Δ-грамматику Γ_1, для которой $h(\Gamma_1) < h(\Gamma)$.

Доказательство теоремы 4.1.

Для исходной А- Δ-грамматики Γ построим эквивалентную А- Δ-грамматику Γ', для которой $S_g(\Gamma') = \langle 0, 0, 0 \rangle$ и $h(\Gamma') \leqslant h(\Gamma)$ (следствие 4.1 и замечание); затем по А- Δ-грамматике Γ' построим эквивалентную А- Δ-грамматику Γ_1, для которой $h(\Gamma_1)$ и $h(\Gamma')$ (следствие 4.2). Тем самым, по А- Δ-грамматике Γ будет построена эквивалентная А- Δ-грамматика Γ_1 такая, что $h(\Gamma_1) < h(\Gamma)$. Применив описанную процедуру к Γ_1, получим эквивалентную А- Δ-грамматику Γ_2 такую, что $h(\Gamma_2 < h(\Gamma_1)$ и т. д. В результате не более чем $(h(\Gamma) - 1)$-кратного применения этой процедуры получим А- Δ-грамматику Γ_0, эквивалентную Γ и такую, что $h(\Gamma_0) \leqslant 1$. Теорема доказана.

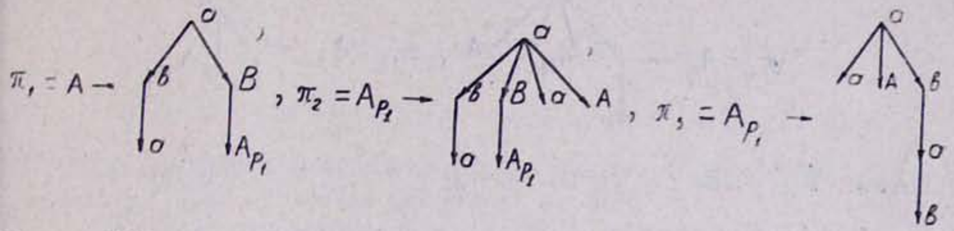
Пример. Пусть задана А- Δ-грамматика $\Gamma = \langle V, W, I, R \rangle$, где $V = \{a, b\}$, $W = \{A, B\}$, $I = A$, $R = \{p_1, p_2, p_3\}$, а правила p_1, p_2 и p_3 заданы схемой 31.



Имеем $W_g^{\Gamma} = \{A\}$, $R_g^{\Gamma} = \{p_1\}$, $S_g(\Gamma) = \langle 3, 1, 1 \rangle$, $S(\Gamma) = \langle 3, 1, 2 \rangle$.

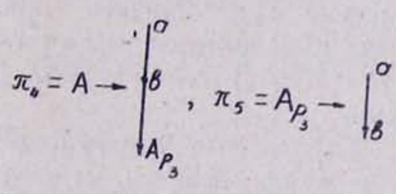
Проделаем несколько шагов по приведению А- Δ-грамматики Γ к нормальной форме.

1) Положим $\Gamma_1 = \langle V, \{A, B, A_{p_1}\}, I, \{p_2, p_3, \pi_1, \pi_2, \pi_3\} \rangle$, где правила π_1, π_2 и π_3 заданы схемой 32.



Имеем $W_g^{\Gamma_1} = \varnothing$, $R_g^{\Gamma_1} = \varnothing$, $S_g(\Gamma_1) = \langle 0, 0, 0 \rangle$, $S(\Gamma_1) = \langle 3, 1, 2 \rangle$.

2) Положим $\Gamma_2 = \langle V, \{A, B, A_{p_1}, A_{p_3}\}, I, \{p_2, \pi_1, \pi_2, \pi_3, \pi_4, \pi_5\} \rangle$, где правила π_4 и π_5 заданы схемой 33.

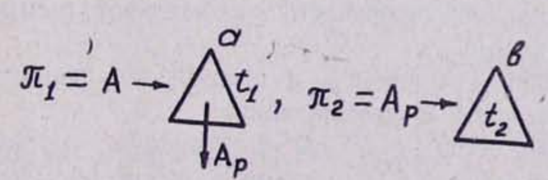


Имеем: $W_g^{\Gamma_2} = \varnothing$, $R_g^{\Gamma_2} = \varnothing$, $S_g(\Gamma_2) = \langle 0, 0, 0 \rangle$, $S(\Gamma_2) = \langle 3, 1, 1 \rangle$.

3) Положим

$$\Gamma_3 = \langle V, \{A, B, A_{p_1}, A_{p_3}, A_{\pi_3}\}, I, \{p_2, \pi_1, \pi_2, \pi_4, \pi_5, \pi_6, \pi_7\} \rangle,$$

где правила π_6 и π_7 заданы схемой 34.

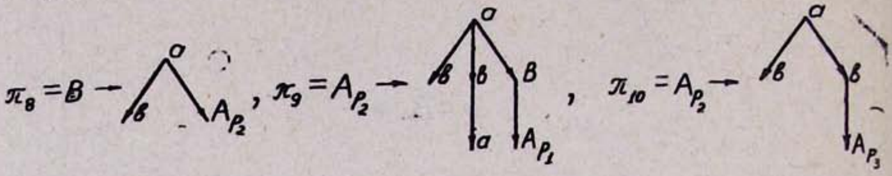


Имеем: $W_g^{\Gamma_3} = \{A, B\}$, $R_g^{\Gamma_3} = \{p_2, \pi_1\}$, $S_g(\Gamma_3) = \langle 2, 1, 2 \rangle$, $S(\Gamma_3) = \langle 2, 2, 1 \rangle$.

4) Положим:

$\Gamma_4 = \langle V, \{A, B, A_{p_1}, A_{p_3}, A_{\pi_1}, A_{p_2}\}, I, \{\pi_1, \pi_3, \pi_4, \pi_5, \ldots, \pi_{10}\} \rangle$,

где правила π_8, π_9, и π_{10} заданы схемой 35.

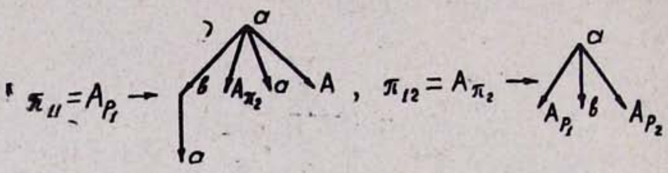


Имеем: $W_g^{\Gamma_4} = \{B\}$, $R_g^{\Gamma_4} = \varnothing$, $S_g(\Gamma_4) = \langle 0, 0, 0 \rangle$, $S(\Gamma_4) = \langle 2, 2, 2 \rangle$.

5) Положим:

$\Gamma_5 = \langle V, \{A, B, A_{p_1}, A_{p_3}, A_{\pi_1}, A_{p_2}, A_{\pi_2}\}, I, \{\pi_1, \pi_4, \pi_5, \ldots, \pi_{12}\} \rangle$,

где правила π_{11} и π_{12} заданы схемой 36.



Имеем: $W_g^{\Gamma_5} = \{B\}$, $R_g^{\Gamma_5} = \varnothing$, $S_g(\Gamma_5) = \langle 0, 0, 0 \rangle$, $S(\Gamma_5) = \langle 2, 2, 1 \rangle$ и т. д.

Дополнение. Квазинормальная форма Л- Δ-грамматики

Пусть $\Gamma = \langle V, W, I, R \rangle$ — линейная Δ-грамматика. Положим $R' = \{p \mid p \in R$ и корень $r(p)$ помечен основным символом$\}$ и $R'' = \{p \mid p \in R$ и корень $r(p)$ помечен вспомогательным символом$\}$.

Таким образом, $R' \cap R'' = \varnothing$ и $R' \cup R'' = R$. Элементы R' назовем автоматными правилами.

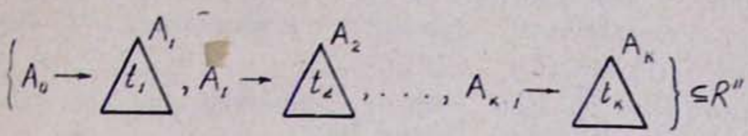
Будем говорить, что линейная Δ-грамматика находится в квазинормальной форме, если для всякого автоматного правила p этой Δ-грамматики $h(p) \leqslant 1$.

Теорема. Д. I. Для каждой линейной Δ-грамматики Γ можно эффективно построить эквивалентную линейную Δ-грамматику в квазинормальной форме.

Таким образом, говоря неформально, «автоматную часть» линейной Δ-грамматики можно привести к нормальной, в смысле автоматных Δ-грамматик, форме.

Доказательству теоремы Д. 1 предпошлем некоторые вспомогательные определения и утверждения.

Пусть $\Gamma = \langle V, W, I, R' \cup R'' \rangle$ — линейная Δ-грамматика. Будем говорить, что символ $B \in W$ *сильно зависит* от символа $A \in W$ в Γ, и писать $A \underset{\Gamma}{\vdash} B$, если либо $A = B$, либо существует последовательность символов из W: $A = A_0, A_1, \ldots, A_k = B$ такая, что для некоторых деревьев $t_1, t_2, \ldots, t_k$ имеет место ситуация, указанная на схеме 37:



Дальнейшие обозначения и определения мало чем отличаются от тех обозначений и определений, которые предваряли доказательство теоремы 4.1.

Для произвольного дерева t обозначим через $g(t)$ множество всех узлов t, высоты которых равны $h(t) - 1$.

Сложностью линейной Δ-грамматики $\Gamma = \langle V, W, I, R' \cup R'' \rangle$ назовем упорядоченную тройку $S'(\Gamma) = \langle h'(\Gamma), M'(\Gamma), N'(\Gamma) \rangle$, где

$$h'(\Gamma) = \max_{p \in R'} h(p),$$

$$M'(\Gamma) = \max_{p \in R',\ h(p) = h'(\Gamma)} \mu(g(r(p))),$$

$$N'(\Gamma) = \mu\{p \mid p \in R',\ h(p) = h'(\Gamma),\ \mu(g(r(p))) = M'(\Gamma)\}).$$

Для произвольных дерева t и множества W обозначим $g_W(t) = \{\alpha \mid \alpha \in g(t)$ и α помечен элементом из $W\}$

Для Л- Δ-грамматики $\Gamma = \langle V, W, I, R' \cup R'' \rangle$ положим:

$$U_\Gamma = \bigcup_{p \in R',\ h(p) = h'(\Gamma)} g_W(r(p)),$$

$$W_g^\Gamma = \{A \mid A \in W \text{ и } A\text{-метка некоторого узла из } U_\Gamma\},$$

$$\widehat{W}_\Gamma = \{B \mid A \vdash_\Gamma B \text{ и } A \in W_g^\Gamma\},$$

$$R_g^\Gamma = \{p \mid p \in R',\ h(p) = h'(\Gamma),\ l(p) \in \widehat{W}_\Gamma \text{ и } g_W(r(p)) \neq \varnothing\}.$$

g-сложностью линейной Δ-грамматики Γ назовем упорядоченную тройку $S_g'(\Gamma) = \langle h_g'(\Gamma), M_g'(\Gamma), N_g'(\Gamma) \rangle$, где

$$h_g'(\Gamma) = \begin{cases} h'(\Gamma), & \text{если } R_g^\Gamma \neq \varnothing, \\ 0, & \text{если } R_g^\Gamma = \varnothing, \end{cases}$$

$$\begin{cases} \max\limits_{p \in R_g^\Gamma} \mu(g_W(r(p))), & \text{если } R_g^\Gamma \neq \varnothing, \\ 0, & \text{если } R_g^\Gamma = \varnothing \end{cases}$$

и

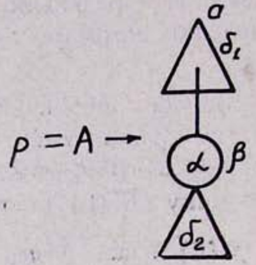
$$N_g'(\Gamma) = \mu(\{p \mid p \in R_g^\Gamma,\ \mu(g_W(r(p))) = M_g'(\Gamma)\}).$$

Отношение $<$ для сложностей S' и g-сложностей S_g' определяется лексикографически.

Ясно, что $S'_g(\Gamma) = \langle 0, 0, 0 \rangle$ тогда и только тогда, когда $R^{\Gamma}_g = \varnothing$.

Лемма Д.1. По всякой линейной Δ-грамматике $\Gamma = \langle V, W, I, R' \cup R'' \rangle$, для которой $S'_g(\Gamma) \neq \langle 0, 0, 0 \rangle$, можно эффективно построить эквивалентную линейную Δ-грамматику Γ' такую, что $S'_g(\Gamma') < S'_g(\Gamma)$ и $h'(\Gamma') \leqslant h'(\Gamma)$.

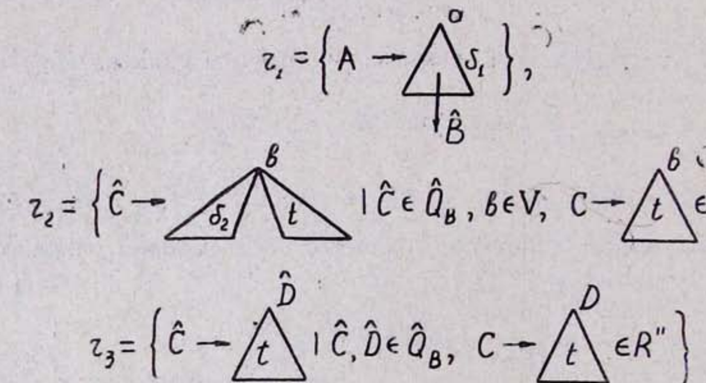
Доказательство. Пусть p-правило из R^{Γ}_g такое, что $\mu(g_W(r(p))) = M'_g(\Gamma)$ (см. схему 38).



где $a \in V$, $A \in \hat{W}_{\Gamma}$, $B \in W^{\Gamma}_g$ $\alpha \in g_W(r(p))$

Для символа B построим множества $Q_B = \{C \mid B \underset{\Gamma}{\vdash} C\}$ и $\hat{Q}_B = \{\hat{C} \mid C \in Q_B\}$ так, чтобы $\hat{Q}_B \cap W = \varnothing$. Нам предстоит рассмотреть два возможных случая: $A \notin Q_B$ и $A \in Q_B$.

1) $A \notin Q_B$. Построим следующие три множества правил (см. схему 39).

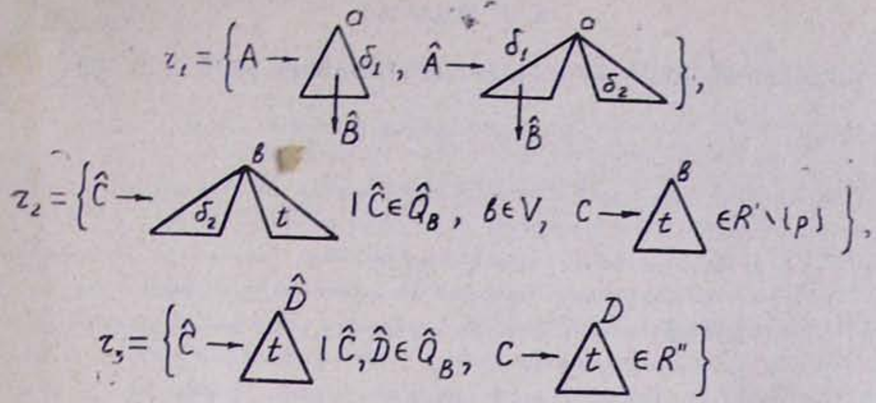


Нетрудно понять, что линейная Δ-грамматика

$$\Gamma' = \langle V, W \cup \hat{Q}_B, I, ((R' \setminus \{p\}) \cup r_1 \cup r_2) \cup (R'' \cup r_3) \rangle$$

эквивалентна Γ и $S'_g(\Gamma') < S'_g(\Gamma)$.

2) $A \in Q_B$. Построим следующие три множества правил (см. схему 40):



Нетрудно убедиться в том, что линейная Δ-грамматика

$$\Gamma' = \langle V,\ W \cup \hat{Q}_B,\ I,\ ((R' \setminus \{p\}) \cup r_1 \cup r_2) \cup (R'' \cup r_3) \rangle$$

эквивалентна Γ, и $S'_g(\Gamma') < S'_g(\Gamma)$.

Заметим, что в обоих случаях $h'(\Gamma') \leqslant h'(\Gamma)$. Лемма доказана.

Следствие Д.1. По всякой линейной Δ-грамматике Γ можно эффективно построить эквивалентную линейную Δ-грамматику Γ', для которой $S'_g(\Gamma') = \langle 0, 0, 0 \rangle$ и $h'(\Gamma') \leqslant h'(\Gamma)$.

Доказательство сводится к многократному применению процедуры леммы Д.1.

Лемма Д.2. По всякой линейной Δ-грамматике Γ, для которой $h'(\Gamma) > 1$ и $S'_g(\Gamma) = \langle 0, 0, 0 \rangle$, можно эффективно построить линейную Δ-грамматику Γ_1 такую, что $S'(\Gamma_1) < S'(\Gamma)$, и, если $h'(\Gamma_1) = h'(\Gamma)$, то $S'_g(\Gamma_1) = \langle 0, 0, 0 \rangle$.

Доказательство не составит большого труда читателю, знакомому с техникой доказательств лемм 4.2 и Д.1; мы его опускаем.

Следствие Д.2. По всякой линейной Δ-грамматике Γ, для которой $h'(\Gamma) > 1$ и $S'_g(\Gamma) = \langle 0, 0, 0 \rangle$, можно эффективно построить эквивалентную линейную Δ-грамматику Γ_1, для которой $h'(\Gamma_1) < h'(\Gamma)$.

Доказательство теоремы Д.1. Для исходной линейной Δ-грамматики Γ построим эквивалентную линейную Δ-грамматику Γ', для которой $S'_g(\Gamma') = \langle 0, 0, 0 \rangle$ и $h'(\Gamma') \leqslant h'(\Gamma)$ (следствие Д.1); затем по Γ' построим эквивалентную линейную Δ-грамматику Γ_1 для которой $h'(\Gamma_1) < h'(\Gamma')$ (следствие Д.2). Тем самым по линейной Δ-грамматике Γ будет построена эквивалентная Δ-грамматика Γ_1, такая, что $h'(\Gamma_1) < h'(\Gamma)$. Остается применить описанную процедуру $h'(\Gamma) - 1$ раз; в результате получим линейную Δ-грамматику в квазинормальной форме, эквивалентную Γ. Теорема доказана.

Ա. Ս. ՍԱՀԱԿՅԱՆ

ԾԱՌԵՐԻ ՔԵՐԱԿԱՆՈՒԹՅՈՒՆՆԵՐԻ ՈՐՈՇ ԴԱՍԵՐԻ ՆՈՐՄԱԼ ՁԵՎԵՐ

Ամփոփում

Առաջարկված աշխատությունում դիտարկվում են ծառերի քերականության (Δ-քերականության) դասակարգման հետ կապված հարցեր. [1] և [2] աշխատություններում, դասական քերականությունների (շղթաների քերականությունների) համանմանությամբ, որոշված են շրջապատից-անկախ և ավտոմատային Δ-քերականություններ, սակայն Δ-քերականությունների դասում շղթաների գծային քերականությունների համար համանմանություն գտնելու խնդիրը մնացել էր չլուծված: Դրված խնդրի լուծման համար, ներկայիս աշխատությունում շրջապատից-անկախ և ավտոմատային Δ-քերականությունների համար սահմանվում է նորմալ ձևի գաղափար և կառուցվում է նշված տիպի Δ-քերականությունը նորմալ ձևին տանող պրոցեդուրա: Այդպիսի մոտեցումը, մի կողմից, թույլ է տալիս ճշտել շրջապատից-անկախ և ավտոմատային Δ-քերականությունների ծառերի ծնման մեխանիզմ, մյուս կողմից, հնարավորություն է տալիս (և նորից շղթաների քերականությունների համանմանությամբ) պարզ և բնական ձևով որոշել գծային Δ-քերականության, դասը, որպես շրջապատից-անկախ և ավտոմատային Δ-քերականությունների միջակա:

ЛИТЕРАТУРА

1. *Гладкий А. В., Мельчук И. А.* Грамматики деревьев. 1. Опыт формализации преобразований синтаксических структур естественного языка. Сб. «Информационные вопросы семиотики, лингвистики и автоматического перевода», вып. 1, 1971, 16—41.
2. *Модина Л. С.* Древесные грамматики и языки. Кибернетика, 1975, № 5, 86—93.