Глава 9

ОПЕРАЦИИ НАД ЯЗЫКАМИ

§ 9.1. Замкнутость относительно элементарных операций

В этой главе мы применяем операции объединения, конкатенации, обращения, замыкания и т.д. к языкам разных типов. Интересно выяснить, какие операции какие классы языков сохраняют, т.е. отображают языки некоторого класса в тот же самый класс. Есть ряд причин интересоваться этим вопросом. Во-первых, знание, сохраняет операция или нет данный класс языков, помогает характеризовать этот класс. Во-вторых, часто бывает легче узнать, что сложный язык относится к некоторому классу, при помощи того факта, что эта принадлежность является результатом различных операций над другими языками в данном классе, чем путем непосредственного конструирования грамматики для этого языка. В-третьих, знание, полученное из изучения операций над языками, может быть использовано при доказательстве теорем, как это было сделано в гл. 7, где мы показали, что класс рекурсивно перечислимых множеств строго содержит рекурсивные множества, используя при доказательстве тот факт, что рекурсивные множества замкнуты относительно дополнения.

Начнем с рассмотрения операций объединения, конкатенации, замыкания Клини и обращения. Используем следующую лемму о “нормальной форме” для контекстно-зависимых языков и языков типа 0:

Лемма 9.1. *Каждый контекстно-зависимый язык порождается контекстно-зависимой грамматикой, в которой все правила имеют форму либо* $\alpha \to \beta$, *где* α *и* β *— цепочки, состоящие из одних только нетерминалов, либо* $A \to b$, *где* A *— нетерминал, а* b *— терминал. Каждый язык типа* 0 *порождается грамматикой типа* 0, *правила которой имеют указанную форму.*

Доказательство. Пусть $G = (V_N, V_T, P, S)$ — контекстно-зависимая грамматика. Каждому $a \in V_T$ сопоставим новый символ X_a. Рассмотрим грамматику $G_1 = (V'_N, V_T, P_1, S)$, где $V'_N = V_N \cup \{X_a \mid a \in V_T\}$. Множество P_1 включает все правила вида $X_a \to a$. Если $\alpha \to \beta \in P$, то $\alpha_1 \to \beta_1 \in P_1$, где цепочки α_1 и β_1 получаются из цепочек α и β путем замещения в них каждого терминала a символом X_a. Доказательство тривиально и мы оставляется его читателю в качестве упражнения. Подобное же доказательство применимо к грамматикам типа 0.

Теорема 9.1 *Классы регулярных*, *контекстно-свободных*, *контекстно-зависимых и рекурсивно перечислимых множеств замкнуты относительно объединения, конкатенации, замыкания и обращения*.

Доказательство. Для класса регулярных множеств доказательство было дано в гл. 3.

Рассмотрим две грамматики: $G_1 = (V_{\mathrm{N}}^{(1)},\ V_{\mathrm{T}}^{(1)}, P_1, S_1)$ и $G_2 = (V_{\mathrm{N}}^{(2)},\ V_{\mathrm{T}}^{(2)}, P_2, S_2)$, причем обе либо контекстно-свободные, либо контекстно-зависимые, либо типа 0. Без потери общности можно предполагать, что $V_{\mathrm{N}}^{(1)} \cap V_{\mathrm{N}}^{(2)} = \varnothing$.

Кроме того, согласно лемме 9.1 и теореме 4.5 можно считать, что правила грамматик G_1 и G_2 имеют форму $\alpha \to \beta$ и $A \to a$, где α и β — цепочки, состоящие из одних только нетерминалов, A — одиночный нетерминал, а a — одиночный терминальный символ. Кроме того, если G_1 и G_2 — контекстно-свободные грамматики, то $\beta = \varepsilon$ подразумевает, что α есть S_1 или S_2 и что α никогда не появляется в правой части никакого правила.

Объединение. Пусть $G_3 = \left(V_{\mathrm{N}}^{(1)} \cup V_{\mathrm{N}}^{(2)} \cup \{S_3\}, V_{\mathrm{T}}^{(1)} \cup V_{\mathrm{T}}^{(2)}, P_3, S_3\right)$, где $S_3 \notin V_{\mathrm{N}}^{(1)} \cup V_{\mathrm{N}}^{(2)}$, а множество P_3 содержит правила $S_3 \to S_1$, $S_3 \to S_2$ и все правила из множеств P_1 и P_2 за исключением $S_1 \to \varepsilon$ и $S_2 \to \varepsilon$, если G_1 и G_2 — контекстно-зависимы. В случае, когда G_1 и G_2 — контекстно-зависимы и $S_1 \to \varepsilon \in P_1$ или $S_2 \to \varepsilon \in P_2$, добавим правило $S_3 \to \varepsilon$ к множеству правил P_3. Теперь грамматика G_3 — того же типа, что и грамматики G_1, G_2, и $L(G_3) = L(G_1) \cup L(G_2)$.

Конкатенация. Пусть $G_4 = (V_{\mathrm{N}}^{(1)} \cup V_{\mathrm{N}}^{(2)} \cup \{S_4\},\ V_{\mathrm{T}}^{(1)} \cup V_{\mathrm{T}}^{(2)}, P_4, S_4)$, где $S_4 \notin V_{\mathrm{N}}^{(1)} \cup V_{\mathrm{N}}^{(2)}$, а множество P_4 содержит правило $S_4 \to S_1S_2$ и все правила из множеств P_1 и P_2, за исключением правил $S_1 \to \varepsilon$ и $S_2 \to \varepsilon$, если G_1 и G_2 — контекстно-зависимы. В случае, когда G_1 и G_2 — контекстно-зависимы и $S_1 \to \varepsilon \in P_1$, добавим правило $S_4 \to S_2$ к множеству правил P_4; если $S_2 \to \varepsilon \in P_2$, то добавим правило $S_4 \to S_1$ к множеству P_4. Если $S_1 \to \varepsilon \in P_1$ и $S_2 \to \varepsilon \in P_2$, то добавим правило $S_4 \to \varepsilon$ к множеству P_4. Теперь грамматика G_4 — того же типа, что и грамматики G_1, G_2 и $L(G_4) = L(G_1)L(G_2)$.

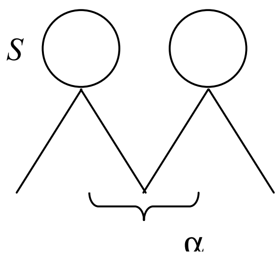

Рис. 9.1.

Заметим, что поскольку $V_{\mathrm{N}}^{(1)} \cap V_{\mathrm{N}}^{(2)} = \varnothing$ и все правила из множеств P_1, P_2 имеют нетерминалы исключительно слева, невозможно, чтобы строка, образованная правым концом сентенциальной формы грамматики G_1, за которой следует левый конец сентенциальной формы грамматики G_2, могла быть левой стороной правила в множестве P_4. Это означает, что левая часть любого правила целиком состоит из нетерминалов только одной из двух грамматик (см. рис. 9.1). Соответственно и все правило относится к одной исходной грамматике. Доказательство того, что $L(G_4) = L(G_1)L(G_2)$ просто.

Замыкание. Пусть $G_5 = (V_N, V_T^{(1)}, P_5, S_5)$, где $V_N = V_N^{(1)} \cup \{S_5, S_5'\}$, а $P_5 = P_1 \cup \{S_5 \to S_1S_5, S_5 \to \varepsilon\}$, если G_5 — контекстно-свободная грамматика, иначе $P_5 = P_1 \cup \{S_5 \to \varepsilon, S_5 \to S_1, S_5 \to S_1 S_5'\} \cup \{a S_5' \to aS_1, a S_5' \to aS_1 S_5' \mid a \in V_T^{(1)}\}$. Однако в случае, когда G_1 — контекстно-зависимая грамматика, правило $S_1 \to \varepsilon$, если оно имеется, отбрасывается. Грамматика G_5 является грамматикой того же типа, что и G_1, и $L(G_5) = (L(G_1))^*$.

Теперь рассмотрим операции *пересечения* и *дополнения*.

Теорема 9.2. *Класс контекстно-свободных языков не замкнут относительно пересечения.*

Доказательство. Языки $L_1 = \{a^n b^n c^i \mid n \geq 1, i \geq 0\}$ и $L_2 = \{a^j b^n c^n \mid n \geq 1, i \geq 0\}$ являются контекстно-свободными, поскольку они порождаются грамматиками $G_1 = (\{S, T\}, \{a, b, c\}, \{S \to Sc, S \to T, T \to aTb, T \to ab\}, S)$ и $G_2 = (\{S, T\}, \{a, b, c\}, \{S \to aS, S \to T, T \to bTc, T \to bc\}, S)$ соответственно. Теперь язык $L = L_1 \cap L_2 = \{a^n b^n c^n \mid n \geq 1\}$, который не контекстно-свободен по тривиальному следствию из теоремы 4.7. Действительно, все цепочки L не соответствуют требуемому условию: в L должна быть цепочка $uvwxy$, такая, что все цепочки вида $uv^n wx^n y$ при любом n тоже принадлежали бы языку L. Мы могли бы считать $u = y = \varepsilon$, но ядро w должно быть в первой степени, а в цепочках из языка L оно тоже в степени n.

Теорема 9.3. *Класс контекстно-свободных языков не замкнут относительно дополнения.*

Доказательство. Поскольку класс контекстно-свободных языков замкнут относительно объединения, но не пересечения, то он не может быть замкнут относительно дополнения, поскольку $L_1 \cap L_2 = \overline{\overline{L_1} \cup \overline{L_2}}$.

Теорема 9.4. *Класс контекстно-свободных языков замкнут относительно пересечения с регулярным множеством.*

Доказательство. Пусть L — контекстно-свободный язык, а R — регулярное множество. Предположим, что $P_1 = (Q_P, \Sigma, \Gamma, \delta_P, p_0, Z_0, F_P)$ — недетерминированный магазинный автомат (npda), принимающий язык L, а $A = (Q_A, \Sigma, \delta_A, q_0, F_A)$ — детерминированный конечный автомат (dfa), принимающий множество R. Построим недетерминированный магазинный автомат (npda) $P_2 = (Q_P \times Q_A, \Sigma, \Gamma, \delta, [p_0, q_0], Z_0, F_P \times F_A)$, который принимает $L \cap R$. Функция δ определяется следующим образом. Для всех $p \in Q_P$, $q \in Q_A$, $a \in \Sigma \cup \{\varepsilon\}$ и $Z \in \Gamma$ функция $\delta([p, q], a, Z)$ содержит $([p', \delta_A(q, a)], \gamma)$ всякий раз, как $\delta_P(p, a, Z)$ содержит (p', γ). (Напомним, что $\delta_A(q, \varepsilon) = q$ для всех $q \in Q_A$.) Неформально npda P_2 хранит след состояний npda P_1 и dfa A в своем конечном управлении.

I. Предположим, что $x \in L \cap R$. Пусть $x = a_1 a_2 \ldots a_n$, где $a_i \in \Sigma \cup \{\varepsilon\}$, $1 \leq i \leq n$, так что существуют состояния $q_0, q_1, \ldots, q_n \in Q_A$, $p_0, p_1, \ldots, p_n \in Q_P$ и цепочки $\gamma_0, \gamma_1, \ldots, \gamma_n \in \Gamma^*$, для которых имеют место $\delta_A(q_i, a_{i+1}) = q_{i+1}$ и $(p_i, a_{i+1} \ldots a_n, \gamma_i) \vdash^{*}_{P_1}$

$(p_{i+1},\ a_{i+2}\ldots a_n,\ \gamma_{i+1})$ при условии, что $0 \le i < n$, $\gamma_0 = Z_0$, $q_n \in F_A$, $p_n \in F_P$. Тогда $([p_i, q_i], a_{i+1}\ldots a_n, \gamma_i) \overset{*}{\underset{P_2}{\vdash}} ([p_{i+1}, q_{i+1}], a_{i+2}\ldots a_n, \gamma_{i+1})$ и $([p_0, q_0], x, Z_0) \overset{*}{\underset{P_2}{\vdash}} ([p_n, q_n], \varepsilon, \gamma_n)$, при том, что $[p_n, q_n] \in F_P \times F_A$, так что $x \in T(P_2)$.

II. Теперь предположим, что $x \in T(P_2)$. Тогда существуют движения вида $([p_i, q_i], a_{i+1}\ldots a_n, \gamma_i) \overset{*}{\underset{P_2}{\vdash}} ([p_{i+1}, q_{i+1}], a_{i+2}\ldots a_n, \gamma_{i+1})$ для $0 \le i < n$, причем $\gamma_0 = Z_0$, $[p_n, q_n] \in F_P \times F_A$. Тогда $\delta_A(q_i, a_{i+1}) = q_{i+1}$ для $0 \le i < n$, причем $q_n \in F_A$. Следовательно, $x \in R$. Аналогично $(p_i, a_{i+1}\ldots a_n, \gamma_i) \underset{P_1}{\vdash} (p_{i+1}, a_{i+2}\ldots a_n, \gamma_{i+1})$ для $0 \le i < n$ и, как следствие, $(p_0, x, Z_0) \overset{*}{\underset{P_1}{\vdash}} (p_n, \varepsilon, \gamma_n)$. Поскольку $p_n \in F_P$, то $x \in L$.

Из рассуждений I и II следует $T(P_2) = L \cap R$.

Мы уже видели в гл. 3, что класс регулярных множеств относительно пересечения и дополнения. В гл. 7 было показано, что класс рекурсивно перечислимых множеств не замкнут относительно дополнения. Таким образом мы имеем:

Теорема 9.5. *Класс языков типа* 0 *не замкнут относительно дополнения.*

В настоящее время неизвестно, замкнут ли класс контекстно-зависимых языков относительно дополнения. Однако как класс языков типа 0, так и класс контекстно зависимых языков замкнуты относительно пересечения. Доказательства для обоих классов аналогичны, и хотя концептуально просты, утомительны в деталях. Поэтому эти доказательства будут только намечены.

Теорема 9.6. *Класс языков типа* 0 *и класс контекстно-зависимых языков замкнуты относительно пересечения.*

Доказательство. Пусть L_1 и L_2 — языки типа 0 (контекстно-зависимые языки). Рассмотрим две одноленточные машины Тьюринга (два недетерминированных линейно ограниченных автомата) M_1 и M_2, принимающие языки L_1 и L_2 соответственно. Легко построить машину Тьюринга (lba) M, имеющую одну оперативную ленту с тремя дорожками. Первая дорожка содержит ввод. Машина M моделирует машину M_1, используя дорожку 2. Если машина M_1 достигает когда-либо принимающую конфигурацию, то машина M перемещает головку своей ленты на левый конец и моделирует машину M_2 на дорожке 3. Если машина M_2 доходит до принимающей конфигурации, то машина M принимает.

§ 9.2. Замкнутость относительно отображений

Теперь рассмотрим результаты отображений разных типов над языками. Первый тип, который мы рассмотрим, — *подстановка*.

***Определение* 9.1.** *Подстановка* f есть отображение конечного множества Σ на подмножества Δ^* некоторого конечного множества Δ. Другими словами, подстановка f с каждым символом из множества Σ ассоциирует некоторый язык. Отображение f может быть распространено на строки из Σ^* следующим образом:

$f(\varepsilon) = \varepsilon$, $f(xa) = f(x)f(a)$, где $x \in \Sigma^*$, $a \in \Sigma$.

Очевидно, что $f(x)$ нужно понимать в обобщенном смысле, тогда как $f(a)$ — в первоначальном.

Мы можем распространить подстановку f далее на языки, определяя $f(L) = \bigcup_{x \in L} f(x)$.

Пример 9.1. Пусть $f(0) = \{a\}$, $f(1) = \{ww^R \mid w \in \{b, c\}^*\}$. Подстановка f отображает множество $\{0^n 1^n \mid n \geq 1\}$ в множество $\{a^n w_1 w_1^R w_2 w_2^R \ldots w_n w_n^R \mid w_i \in \{b, c\}^*$ для $1 \leq i \leq n\}$.

Говорят, что класс языков *замкнут относительно подстановки*, если для любого языка $L \subseteq \Sigma^*$ в данном классе и для любой подстановки f, такой, что $f(a)$ в данном классе для всех $a \in \Sigma$, язык $f(L)$ содержится в этом же классе.

Покажем, что классы регулярных множеств, контекстно-свободных языков и языков типа 0 замкнуты относительно подстановки. Так, в примере 9.1, поскольку $f(0)$ и $f(1)$ — оба контекстно-свободные языки и так как $L = \{0^n 1^n \mid n \geq 1\}$ — контекстно-свободный язык, то множество $f(L) = \{a^n w_1 w_1^R w_2 w_2^R \ldots w_n w_n^R \mid w_i \in \{b, c\}^*$ для $1 \leq i \leq n\}$ также контекстно-свободный язык.

Теорема 9.7. *Классы регулярных множеств, контекстно-свободных языков и языков типа* 0 *замкнуты относительно подстановки.*

Доказательство. Рассмотрим грамматику $G = (V_N, \{a_1, a_2, \ldots, a_n\}, P, S)$. Пусть $G_i = (V_{N_i}, V_{T_i}, P_i, S_i)$ — грамматика, порождающая множество $f(a_i)$ для каждого i, $1 \leq i \leq n$. Без потери общности предполагаем, что все нетерминальные словари попарно не пересекаются.

Докажем теорему для случая, когда грамматики G и G_i, $1 \leq i \leq n$, являются контекстно-свободными. Читатель может доказать другие случаи аналогично, хотя в каждом необходимы дополнительные детали.

Построим новую грамматику:

$$G' = (V_N', V_T', P', S), \text{ где } V_N' = V_N \cup \bigcup_{i=1}^{n} V_{N_i},\ V_T' = \bigcup_{i=1}^{n} V_{T_i}.$$

Пусть h — подстановка $h(a_i) = \{S_i\}$ для $1 \leq i \leq n$ и $h(A) = \{A\}$ для любого $A \in V_N$; $P' = \bigcup_{i=1}^{n} P_i \cup \{A \to h(\alpha) \mid A \to \alpha \in P\}$. Ясно, что грамматика G' является контекстно-свободной, возможно, с правилами вида $A \to \varepsilon$. Очевидно, что $f(L(G)) = L(G')$.

Пример 9.2. Пусть $L = \{0^n 1^n \mid n \geq 1\}$. Язык L порождается грамматикой

$$G = (\{S\}, \{0, 1\}, \{S \to 0S1, S \to 01\}, S).$$

Как и в примере 9.1, пусть

$$f(0) = \{a\} \text{ и } f(1) = \{ww^R \mid w \in \{b, c\}^*\};$$

$f(0)$ порождается грамматикой

$$G_1 = (\{S_1\}, \{a\}, \{S \to a\}, S_1),$$

а $f(1)$ — грамматикой

$$G_2 = (\{S_2\},\ \{\mathrm{b},\mathrm{c}\},\ \{S_2 \to bS_2b,\ S_2 \to cS_2c,\ S_2 \to \varepsilon\},\ S_2).$$

Язык $f(L)$ порождается грамматикой $G_3 = (\{S, S_1, S_2\}, \{a, b, c\}, \{S \to S_1SS_2,\ S \to S_1S_2,$ $S_1 \to a,\ \ S_2 \to bS_2b,\ S_2 \to cS_2c,\ S_2 \to \varepsilon\},\ S)$. Первые два правила грамматики G_3 получились из правил $S \to 0S1$ и $S \to 01$ грамматики G_1 в результате подстановки символа S_1 вместо 0 и символа S_2 — вместо 1.

Контекстно-зависимые языки не замкнуты относительно подстановки. Однако мы можем несколько смягчить этот факт.

***Определение* 9.2.** Подстановка f называется *ε-свободной* (ε-free), если $\varepsilon \notin f(a)$ для каждого $a \in \Sigma$.

Теорема 9.8. *Класс контекстно-зависимых языков замкнут относительно ε-свободной подстановки.*

Доказательство. Рассмотрим контекстно-зависимую грамматику $G = (V_{\mathrm{N}},$ $\{a_1,\ a_2, \ldots,\ a_n\},\ P,\ S)$ и ε-свободную подстановку f. Пусть для каждого i, $1 \le i \le n$, $G_i = (V_{\mathrm{N}_i}, V_{\mathrm{T}_i}, P_i, S_i)$ — контекстно-зависимая грамматика, порождающая множество $f(a_i)$. Без потери общности предполагаем, что все нетерминальные словари попарно не пересекаются. Кроме того, предполагаем, что все правила, за возможным исключением $S \to \varepsilon$, имеют вид $\alpha \to \beta$ или $A \to a$, где α, β — непустые строки нетерминалов, A — отдельный нетерминал, а a — отдельный терминальный символ. Мы построим грамматику $G' = (V_{\mathrm{N}}', V_{\mathrm{T}}', P', S_L)$, где

1. $V_{\mathrm{N}}' = V_{\mathrm{N}} \cup \bigcup_{i=1}^{n} V_{\mathrm{N}_i} \cup \{A_L \mid A \in V_{\mathrm{N}}\}$.

2. $V_{\mathrm{T}}' = \bigcup_{i=1}^{n} V_{\mathrm{T}_i}$.

3. P' содержит

а) $S_L \to \varepsilon$, если $S \to \varepsilon \in P$;

б) $A_L\alpha \to B_L\beta$ и $A\alpha \to B\beta$, если $A\alpha \to B\beta \in P$ (заметим, что индекс L в обозначении A_L помечает самое левое вхождение соответствующего нетерминального символа в выводе в грамматике G до тех пор, пока этот символ не превратится в терминальный символ);

в) $A_L \to S_i$, если $A \to a_i \in P$,

$aA \to aS_i$ для всех $a \in V_{\mathrm{T}}'$, если $A \to a_i \in P$;

г) все правила из множества $\{P_i \mid i = 1, 2, \ldots, n\}$.

Грамматика G' — контекстно-зависимая и $L(G') = f(L(G))$.

Теорема 9.9. *Класс контекстно-зависимых языков не замкнут относительно подстановки.*

Доказательство. Пусть $G_1 = (V_{\mathrm{N}}, V_{\mathrm{T}}, P_1, S)$ — грамматика типа 0, такая, что $L(G_1)$ не является контекстно-зависимым языком. Снова мы предполагаем без потери общности, что все ее правила имеют вид $\alpha \to \beta$ или $A \to a$, где $\alpha \in V_{\mathrm{N}}^{+}$, $\beta \in V_{\mathrm{N}}^{*}$, $A \in V_{\mathrm{N}}$, $a \in V_{\mathrm{T}}$.

Пусть c — новый символ. Рассмотрим грамматику $G_2 = (V_N, V_T \cup \{c\}, P_2, S)$, в которой P_2 содержит

1) $\alpha \to \beta$, если $\alpha \to \beta \in P_1$ и $|\alpha| \le |\beta|$;

2) $\alpha \to \beta cc\ldots c$, где $|\alpha| = |\beta cc\ldots c|$, если $\alpha \to \beta \in P_1$ и $|\alpha| > |\beta|$;

3) $cA \to Ac$ для всех $A \in V_N$.

Грамматика G_2 является контекстно-зависимой, поскольку мы принудили правую часть каждого правила иметь, по крайней мере, такую же длину, как левая. Правила $cA \to Ac$ были добавлены для того, чтобы передвигать символы c к правому концу слов так, чтобы выводы в G_2 могли происходить, как в грамматике G_1.

Теперь рассмотрим подстановку $f(a) = \{a\}$ для $a \in V_T$ и $f(c) = \{\varepsilon\}$. Тогда $f(L(G_2)) = L(G_1)$ и, следовательно, подстановка не сохраняет класс csl.

Очень часто интерес представляют подстановки специальных типов.

***Определение* 9.3.** Подстановка f называется *конечной*, если $f(a)$ есть конечное множество для всех a из области определения f. Если $f(a)$ — единственная строка, то f — *гомоморфизм*.

Конечная подстановка и гомоморфизм являются специальными классами подстановок. Из этого мы имеем следующие следствия:

***Следствие* 9.1.** Классы регулярных, контекстно-свободных и языков типа 0 замкнуты относительно конечной подстановки и гомоморфизма.

Доказательство очевидно из теоремы 9.7.

***Следствие* 9.2.** Класс контекстно-зависимых языков замкнут относительно ε-свободной конечной подстановки и ε-свободного гомоморфизма.

Доказательство очевидно из теоремы 9.8.

***Следствие* 9.3.** Класс контекстно-зависимых языков не замкнут относительно конечной подстановки и гомоморфизма.

Доказательство. Подстановка, использованная при доказательстве теоремы 9.9, является гомоморфизмом.

Мы докажем еще один результат, касающийся подстановок, поскольку он необходим для последующей теоремы.

***Определение* 9.4.** Класс языков замкнут относительно *k-ограниченного стирания*, если для любого языка L этого класса и любого гомоморфизма h, обладающего тем свойством, что h никогда не отображает более, чем k последовательных символов любого предложения из языка L в ε, $h(L)$ находится в этом же классе.

Покажем, что класс контекстно-зависимых языков замкнут относительно k-ограниченного стирания. Фактически справедливо более общее утверждение. Пусть $L \subseteq \Sigma^*$ — контекстно-зависимый язык и пусть $f(a)$ для любого $a \in \Sigma$ тоже контекстно-зависим. Тогда язык $f(L)$ контекстно-зависим при условии, что существует $k > 0$, такое, что для $x \in L$ и $y \in f(x)$ выполняется неравенство $|y| \ge k|x|$.

Лемма 9.2. *Класс контекстно-зависимых языков замкнут относительно k-ограниченного стирания.*

Доказательство. Пусть $G_1 = (V_N^{(1)},\ V_T^{(1)},\ P_1,\ S_1)$ — контекстно-зависимая грамматика. Без потери общности предположим, что правила, за возможным исключением $S_1 \to \varepsilon$, имеют вид $\alpha \to \beta$ или $A \to a$, где $\alpha, \beta \in V_N^{(1)+}$, $A \in V_N^{(1)}$, а $a \in V_T^{(1)}$. Пусть h — гомоморфизм со свойством, что h никогда не отображает более, чем k последовательных символов любого предложения $x \in L(G_1)$ в ε. Пусть целое l больше, чем $k+1$, и больше длины самой длинной левой части любого правила. Рассмотрим грамматику

$$G_2 = \left(V_N^{(2)}, V_T^{(2)}, P_2, S_2\right),$$

где

$$V_N^{(2)} = \left\{[\alpha] \mid \alpha \in (V_N^{(1)} \cup V_T^{(1)})^*, |\alpha| < 2l\right\},$$

$V_T^{(2)}$ содержит такие символы, находящиеся в строках w, что $h(a) = w$ для некоторого $a \in V_T^{(1)}$, $S_2 = [S_1]$, а множество правил P_2 содержит

1) $[S_1] \to \varepsilon$, если $S_1 \to \varepsilon \in P_1$ или если $x \in L(G_1)$ и $h(x) = \varepsilon$ (заметим, что $|x| \le k$, так что мы можем проверить, существует ли какая-нибудь такая цепочка x);

2) $[\alpha] \to [\beta]$ для всех $[\alpha]$ и $[\beta]$ из $V_N^{(2)}$, таких, что $\alpha \underset{G_1}{\Rightarrow} \beta$ и $|\beta| < 2l$;

3) $[\alpha] \to [\beta_1][\beta_2]\ldots[\beta_m]$ для всех $[\alpha]$, $[\beta_1]$, $[\beta_2], \ldots$, $[\beta_m]$ из $V_N^{(2)}$, таких, что $\alpha \underset{G_1}{\Rightarrow} \beta_1\beta_2\ldots\beta_m$, $|\beta_i| = l$, $1 \le i < m$, $l \le |\beta_m| < 2l$;

4) $[\alpha_1][\alpha_2] \to [\beta_1][\beta_2]\ldots[\beta_m]$ для всех $[\alpha_1]$, $[\alpha_2]$, $[\beta_1]$, $[\beta_2], \ldots$, $[\beta_m]$ из $V_N^{(2)}$, таких, что $\alpha_1\alpha_2 \underset{G_1}{\Rightarrow} \beta_1\beta_2\ldots\beta_m$, $l \le |\alpha_1| < 2l$, $l \le |\alpha_2| < 2l$, $|\beta_i| = l$, $1 \le i < m$, $l \le |\beta_m| < 2l$;

5) $[x] \to h(x)$ для всех $[x] \in V_N^{(2)}$, $x \in V_T^{(1)*}$, $h(x) \ne \varepsilon$.

Грамматика G_2 является контекстно-зависимой и $L(G_2) = h(L(G_1))$. Отметим, что G_2 получается путем кодирования блоков по меньшей мере из $k+1$ символа грамматики G_1 в один символ. Поскольку не более k последовательных терминальных символов грамматики G_1 отображаются в ε, то в грамматике G_2 никогда не требуется иметь правило, в котором нетерминал, не равный начальному, порождал бы ε.

***Определение* 9.5.** *Обобщенная последовательная машина* (gsm[10]) есть конечный автомат, который может выводить конечное число символов для каждого входного символа. Формально обобщенная последовательная машина есть система $S = (Q, \Sigma, \Delta, \delta, q_0, F)$, где Q — *состояния*; Σ — *входной алфавит*; Δ — *выходной алфавит*; δ — *отображение типа* $Q \times \Sigma \to 2^{Q \times \Delta^*}$; $q_0 \in Q$ — *начальное*

[10] Gsm — generalized sequential machine. В качестве синонима в настоящее время используется более современный термин *конечный преобразователь* (finite transducer — ft).

состояние; $F \subseteq Q$ — *множество конечных состояний*. Запись $(p, w) \in \delta(q, a)$ означает, что S в состоянии q, имея на входе символ a, может в качестве одного из возможных вариантов движения перейти в состояние p и вывести строку w.

Мы расширим область определения δ до $Q \times \Sigma^*$ следующим образом:

$\delta(q, \varepsilon) = \{(q, \varepsilon)\}$,

$\delta(q, xa) = \{(p, w) \mid w = w_1 w_2,\ (p', w_1) \in \delta(q, x)$ и $(p, w_2) \in \delta(p', a)\}$, если $x \in \Sigma^*$ и $a \in \Sigma$.

***Определение* 9.6.** Пусть S — обобщенная последовательная машина и $S(x) = \{y \mid (p, y) \in \delta(q_0, x)$ для некоторого $p \in F\}$. Если L есть язык над Σ, то $S(L) = \{y \mid y \in S(x)$ для некоторого $x \in L\}$ называется *gsm-отображением*, а $S^{-1}(L) = \{y \mid x \in S(y)$ для некоторого $x \in L\}$ — *обратным gsm-отображением*.

Не обязательно истинно, что $S^{-1}(S(L)) = S(S^{-1}(L)) = L$, и потому отображение S^{-1} не является подлинно обратным

Пример 9.3. Пусть $S = (\{q_0, q_1\}, \{0, 1\}, \{a, b\}, \delta, q_0, \{q_1\})$ — обобщенная последовательная машина, где отображение δ определено следующим образом:

1) $\delta(q_0, 0) = \{(q_0, aa), (q_1, b)\}$,
2) $\delta(q_0, 1) = \{(q_0, a)\}$,
3) $\delta(q_1, 0) = \varnothing$,
4) $\delta(q_1, 1) = \{(q_1, \varepsilon)\}$.

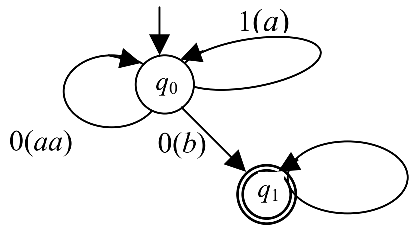

Рис. 9.2.

Интуитивно, пока в gsm S вводятся нули (рис. 9.2), gsm S имеет выбор: выводить два символа a либо один символ b. Если gsm S выводит b, она переходит в состояние q_1. Если gsm S находится в состоянии q_0 и в нее вводится символ 1, то она может выводить только символ a. В состоянии q_1 gsm S ничего не может поделать с 0 на входе, но может оставаться в состоянии q_1 без какого-либо вывода, если на входе 1.

Пусть $L = \{0^n 1^n \mid n \geq 1\}$. Тогда $S(L) = \{a^{2n} b \mid n \geq 0\}$.

Если обозначить $S(L)$ при помощи L_1, то $S^{-1}(L_1) = \{w01^i \mid i \geq 0$ и w имеет[11] четное число 1$\}$. Заметим, что $S^{-1}(S(L)) \neq L$.

Характерная особенность gsm-отображения и обратного gsm-отображения состоит в том, что они сохраняют разные классы языков.

[11] При этом w может содержать любое число нулей.

Лемма 9.3. *Каждый класс языков, замкнутый относительно конечной подстановки и пересечения с регулярным множеством, замкнут относительно* gsm-*отображений.*

Доказательство. Пусть C — класс языков, замкнутый относительно конечной подстановки (следовательно, также и гомоморфизма) и пересечения с регулярным множеством. Пусть $S = (Q, \Sigma, \Delta, \delta, q_0, F)$ — обобщенная последовательная машина. Определим конечную подстановку

$f(a) = \{[q, a, x, p] \mid q, p \in Q, a \in \Sigma, x \in \Delta^*, \text{и } (p, x) \in \delta(q, a)\}$.

Пусть R — регулярное множество, содержащее все строки вида

$$[q_0, a_1, x_1, q_1]\,[q_1, a_2, x_2, q_2] \ldots [q_{n-1}, a_n, x_n, q_n],$$

такие, что для $1 \le i \le n$, $a_i \in \Sigma$, $x_i \in \Delta^*$, $q_i \in Q$, $(q_i, x_i) \in \delta(q_{i-1}, a_i)$. Также q_0 — начальное состояние и $q_n \in F$. Пусть $h([q, a, x, p]) = x$ для всех $[q, a, x, p]$.

Теперь для $L \in C$ имеем $S(L) = h(f(L) \cap R)$. Поскольку класс языков C замкнут относительно конечной подстановки и пересечения с регулярным множеством, то язык $S(L)$ тоже находится в C. Заметим, что требуется замкнутость относительно конечной подстановки, а не ε-свободной конечной подстановки, поскольку в $[q, a, x, p]$ цепочка x может быть равна ε, и в этом случае $h([q, a, x, p]) = \varepsilon$.

Теорема 9.10. *Классы регулярных, контекстно-свободных и языков типа 0 замкнуты относительно gsm-отображений.*

Доказательство. Теорема является прямым следствием леммы 9.3 и теорем 9.4, 9.6 и 9.7.

Отметим, что gsm-отображения не сохраняют контекстно-зависимых языков, поскольку каждый гомоморфизм является gsm-отображением.

***Определение* 9.7.** Говорят, что gsm-отображение ε-*свободно*, если $(p, \varepsilon) \notin \delta(q, a)$ для любых $q, p \in Q$ и $a \in \Sigma$.

Хотя контекстно-зависимые языки не замкнуты относительно произвольных gsm-отображений, они замкнуты относительно ε-свободных gsm-отображений.

Теорема 9.11. *Класс контекстно-зависимых языков замкнут относительно ε-свободных gsm-отображений.*

Доказательство. В лемме 9.3 конечная подстановка может быть заменена на ε-свободную конечную подстановку при условии, что gsm-отображение ε-свободно. Таким образом, поскольку класс контекстно-зависимых языков замкнут относительно ε-свободной конечной подстановки и пересечения с регулярным множеством, то этот класс замкнут относительно ε-свободных gsm-отображений.

Рассмотрим теперь обратные gsm-отображения. Как увидим, регулярные, контекстно-свободные, контекстно-зависимые и языки типа 0 все замкнуты относительно обратных gsm-отображений.

Лемма 9.4. *Пусть C — класс языков, замкнутый относительно ε-свободной подстановки, k-ограниченного стирания и объединения и пересечения с регулярными множествами. Тогда класс C замкнут относительно обратных gsm-отображений.*

Доказательство. Пусть $L \subseteq \Delta^*$ есть язык в классе C, а $S = (Q, \Sigma, \Delta, \delta, q_0, F)$ — обобщенная последовательная машина. Мы предполагаем без потери общности, что $\Sigma \cap \Delta = \varnothing$. Определим подстановку f следующим образом: $f(b) = \Sigma^* b$ для каждого $b \in \Delta$. (Отметим, что замкнутость относительно объединения и пересечения с регулярными множествами гарантирует принадлежность всех регулярных множеств классу C и, следовательно, $\Sigma^* b \in C$.)

Пусть $L_1 = f(L) \cup \Sigma^*$, если $\varepsilon \in L$, и $L_1 = f(L)$ в противном случае. Тогда L есть множество всех строк вида $y_1 b_1 y_2 b_2 \ldots y_r b_r$, $r \geq 1$, где $b_i \in \Delta$, $y_i \in \Sigma^*$, $1 \leq i \leq r$, $b_1 b_2 \ldots b_r \in L$, объединенное с Σ^*, если $\varepsilon \in L$. Применим теперь лемму 9.3 к классам регулярных, контекстно-свободных и языков типа 0.

Пусть R — регулярное множество, состоящее из всех слов вида $a_1 x_1 a_2 x_2 \ldots a_m x_m$, $m \geq 0$, таких, что

1) $a_i \in \Sigma$;

2) $x_i \in \Delta^*$, $1 \leq i \leq m$.

Существуют состояния $q_0, q_1, \ldots, q_m$, такие, что $q_m \in F$ и $(q_i, x_i) \in \delta(q_{i-1}, a_i)$ для $1 \leq i \leq m$.

Заметим, что цепочка x_i может быть равна ε. Нетрудно показать путем построения конечного автомата, принимающего R, что R — регулярное множество.

Теперь $L_1 \cap R$ есть множество всех слов вида $a_1 x_1 a_2 x_2 \ldots a_m x_m$, $m \geq 0$, где $a_i \in \Sigma$, $x_i \in \Delta^*$, $1 \leq i \leq m$, $x_1 x_2 \ldots x_m \in L$, $S(a_1 a_2 \ldots a_m)$ содержит цепочку $x_1 x_2 \ldots x_m$, и ни одна цепочка x_i не длиннее, чем k, причем k — длина самой длинной цепочки x, такой, что $(p, x) \in \delta(q, a)$ для некоторых состояний $p, q \in Q$ и $a \in \Sigma$.

Наконец, пусть h — гомоморфизм, который отображает символ a в a для каждого $a \in \Sigma$ и символ b — в ε для каждого $b \in \Delta$. Тогда $S^{-1}(L) = h(L_1 \cap R)$ находится в классе C, поскольку h никогда не отображает больше k последовательных символов в ε.

Теорема 9.12. Классы регулярных, контекстно-свободных, контекстно-зависимых и языков типа 0 замкнуты относительно обратных gsm-отображений.

Доказательство следует непосредственно из леммы 9.4 и того факта, что названные классы замкнуты относительно ε-свободной подстановки, k-ограниченного стирания, а также пересечения и объединения с регулярным множеством.

Теперь рассмотрим операцию деления.

***Определение* 9.8.** Пусть L_1 и L_2 — любые два языка. Определим *частное от деления L_1 на L_2* как множество $\{x \mid$ для некоторой цепочки $y \in L_2$, такой, чтобы $xy \in L_1\}$.

Пример 9.4. Пусть $L_1 = \{a^n b^n \mid n \geq 1\}$ и $L_2 = b^*$. Тогда

$$L_1 / L_2 = \{a^i b^j \mid i \geq j,\, i \geq 1\},\ \text{а}\ L_2 / L_1 = \varnothing.$$

Лемма 9.5. *Каждый класс языков, замкнутый относительно конечной подстановки и пересечения с регулярным множеством, замкнут относительно деления на регулярное множество.*

Доказательство. Пусть C — класс языков, замкнутый относительно названных операций. Пусть $L \in \Sigma_1^*$ — язык из класса C и $R \subseteq \Sigma_1^*$ — регулярное множество. Пусть $\Sigma_2 = \{a' \mid a \in \Sigma_1\}$ и f — конечная подстановка: $f(a) = \{a, a'\}$. Рассмотрим $L_2 = \Sigma_2^* R \cap f(L)$. Пусть h — гомоморфизм, определяемый следующим образом: $h(a) = \varepsilon$ и $h(a') = a$ для всех $a \in \Sigma_1$. Теперь $L / R = h(L_2)$. Поскольку класс C замкнут относительно конечной подстановки и пересечения с регулярным множеством, то L / R находится в классе C.

Теорема 9.13. *Классы регулярных, контекстно-свободных и языков типа 0 замкнуты относительно деления на регулярное множество.*

Доказательство следует непосредственно из леммы 9.5.

На вопрос: замкнут ли класс контекстно-зависимых языков относительно деления на регулярное множество, ответим — нет.

Теорема 9.14. *Если L_1 есть любой язык типа* 0, *то существует контекстно-зависимый язык L_2 и регулярное множество R, такие, что $L_1 = L_2 / R$.*

Доказательство почти идентично доказательству теоремы 9.9. Пусть $G_1 = (V_N, V_T, P_1, S_1)$ — грамматика типа 0, такая, что $L(G_1) = L_1$ и пусть $G_2 = (V_N \cup \{S_1, D\}, V_T \cup \{c, d\}, P_2, S_2)$, где P_2 определяется следующим образом:

1) если $\alpha \to \beta \in P_1$ и $|\alpha| \leq |\beta|$, то $\alpha \to \beta \in P_2$;

2) если $\alpha \to \beta \in P_1$ и $|\alpha| - |\beta| = i$, $i > 0$, то $\alpha \to \beta D^i \in P_2$;

3) для всех $A \in V_N$ и $a \in V_T$ существуют правила $DA \to AD$ и $Da \to aD \in P_2$;

4) существуют правила $Dc \to cc$ и $Dc \to dc \in P_2$;

5) существует правило $S_2 \to S_1 Dc \in P_2$.

Обратим внимание читателя на сходство $L(G_2)$ с языком, определенным в теореме 9.9. Но здесь мы можем превращать все нетерминалы D в терминальные символы, если только они сначала мигрируют к правому концу сентенциальной формы. Причем как только нетерминал D превращается в терминал d, ни один символ D больше не может быть превращен ни в d, ни в c. Теорема следует из наблюдения, что $L(G_1) = L(G_2) / dc^*$.

В заключение главы приведем сводку описанных в ней свойств замкнутости для регулярных, контекстно-свободных, контекстно-зависимых и языков типа 0, — см. табл. 9.1.

Табл. 9.1

Замкнутость относительно операций	Класс языка			
	рег.	конт.-свобод.	конт.-завис.	типа 0
Объединение	+	+	+	+
Конкатенация	+	+	+	+
Замыкание	+	+	+	+
Обращение	+	+	+	+
Пересечение	+	–	+	+
Дополнение	+	–	?	–
Пересечение с регулярным множеством	+	+	+	+
Подстановка	+	+	+	+
ε-Свободная подстановка	+	+	+	+
gsm-Отображение	+	+	+	+
ε-Свободное gsm-отображение	+	+	+	+
Обратные gsm-отображения	+	+	+	+
k-Ограниченное стирание	+	+	+	+
Деление на регулярное множество	+	+	+	+

В табл. 9.1 символом + отмечен тот факт, что соответствующий класс языков обладает свойством замкнутости относительно соответствующей операции; символ – (минус) обозначает отсутствие соответствующего свойства замкнутости для соответствующего класса языков; символ ? (вопросительный знак) означает, что пока не выяснено, замкнут ли класс контекстно-зависимых языков относительно дополнения или не замкнут.