

Основы теории формальных языков и грамматик

Глава 1

ГЛАВА 1

Основы теории формальных языков и грамматик

- 1.1 Языки и цепочки символов. Способы задания языков
 - 1.1.1 Цепочки символов. Операции над ними
 - 1.1.2 Формальное определение языка. Понятие языка
 - 1.1.3 Способы задания языка
 - 1.1.4 Синтаксис и семантика
- 1.2 Определение грамматики
 - 1.2.1 Понятие о грамматике языка
 - 1.2.2 Формальное определение грамматики
- 1.3 Способы записи синтаксиса языка
 - 1.3.1 Метаязык Хомского
 - 1.3.2 Бэкуса-Наура формы (БНФ)
 - 1.3.3 РБНФ (расширенная)
 - 1.3.4 Диаграмма Вирта
- 1.4 Классификация языков и грамматик
 - 1.4.1 Классификация грамматик
 - 1.4.2 Классификация языков
 - 1.4.3 Примеры классификации языков и грамматик
- 1.5 Цепочки вывода. Сентенциальная форма
 - 1.5.1 Вывод. Цепочка вывода.
 - 1.5.2 Сентенциальная форма грамматики. Основа
 - 1.5.3 Левосторонний и правосторонний вывод
 - 1.5.4 Дерево вывода

ГЛАВА 1. Основы теории формальных языков и грамматик

1.1 ЯЗЫКИ И ЦЕПОЧКИ СИМВОЛОВ. **СПОСОБЫ ЗАДАНИЯ ЯЗЫКОВ**

1.1.1 Цепочки символов. Операции над ними

- **Цепочкой (строкой)** называется последовательность символов записанных один за одним. $\alpha \beta \gamma \omega$
- Цепочка – последовательность, в которую могут входить все допустимые символы (не обязательно несущие смысл). abc или call_me_1_02
- Цепочки символов α и β равны ($\alpha = \beta$) тогда и только тогда, когда имеют один и тот же состав символов, и одинаковое их количество и их порядок следования.
- Количество символов в цепочке называется длиной цепочки. $|\alpha|$
 $\alpha = abc \quad |\alpha| = 3$
 $\alpha = \beta \quad |\alpha| = |\beta|$

1.1.1 Цепочки символов. Операции над ними

- Если из цепочки единичной длины $|\alpha|=1$ удаляется этот единственный символ, α по прежнему остается цепочкой, но длина ее равна 0. $|\alpha|=0$
- Цепочку нулевой длины будем обозначать ε .

$$|\varepsilon|=0$$

$$\varepsilon d = d\varepsilon$$

- Если существует цепочка $\omega = \alpha\beta$, то α – голова цепочки, β – хвост цепочки.
- Причем α – правильная голова, если β – не пустая цепочка. $|\beta| > 0$.
 β – правильный хвост, если α – не пустая цепочка. $|\alpha| > 0$.

$$\alpha = abc$$

ε, a, ab, abc – головы цепочки. ε, a, ab – правильные головы.

1.1.1 Цепочки символов. Операции над ними

- Если α и β - цепочки, то цепочка $\alpha\beta$ называется **конкатенацией** (или сцеплением) цепочек α и β .

$$\alpha = ab \text{ и } \beta = cd, \alpha\beta = abcd.$$

$$\alpha\varepsilon = \varepsilon\alpha = \alpha.$$

- Коммутативность конкатенации $\alpha\beta \neq \beta\alpha$, ассоциативность $\alpha(\beta\gamma) = (\alpha\beta)\gamma$
- Обращением (или реверсом)** цепочки α называется цепочка, символы которой записаны в обратном порядке. α^R .

$$\alpha = abcdef, \alpha^R = fedcba.$$

$$\varepsilon = \varepsilon^R.$$

$$(\alpha\beta)^R = \beta^R\alpha^R$$

- Итерация (повторение, степень)** n -ой степенью цепочки α (будем обозначать α^n) называется конкатенация n цепочек α .

$$\alpha^0 = \varepsilon; \alpha^n = \alpha\alpha^{n-1} = \alpha^{n-1}\alpha.$$

$$\varepsilon^n = \varepsilon, \text{ где } n \in \mathbb{N}, n \geq 0.$$

1.1.2 Формальное определение языка.

Понятие языка

- **Язык** – это заданный набор символов и правил, устанавливающих способы комбинации этих символов между собой для записи осмысленных предложений.
- **Алфавит** – набор допустимых символов языка. Алфавит – счетное, непустое множество символов.
- Цепочка символов α является цепочкой над алфавитом $\alpha(V)$, если в нее входят только символы, принадлежащие алфавиту V .
- Для любого алфавита V пустая цепочка ε может как являться, так и не являться цепочкой над этим алфавитом.
- Если $|V|=0$ и V – множество, то оно называется пустым множеством и обозначается \emptyset .

$$|\varepsilon|=0$$

$$|\{\varepsilon\}|=1$$

1.1.2 Формальное определение языка.

Понятие языка

- V^* множество, содержащее все цепочки в алфавите V , включая пустую цепочку ϵ .
- V^* - **итерация** множества V или транзитивное замыкание.
- V^+ - множество всех цепочек длиной 1 и более, исключив тем самым цепочку ϵ .
- V^+ - **усечённая итерация** множества V или усеченное транзитивное замыкание.

$$V^* = V^+ \cup \{\epsilon\}$$

$$V = \{a, b, c\}$$

$$V^* = \{a, b, c, aa, bb, cc, aab, abc, abbc \dots \epsilon\}$$

$$V^+ = \{a, b, c, aa, bb, cc, aab, abc, abbc \dots\}$$

- **Декартовым произведением** $A \times B$ множеств A и B называется множество $\{\alpha \beta \mid \alpha \in A, \beta \in B\}$.

$$\text{Если } A = \{a, b\} \text{ и } B = \{c, d\}, \text{ то } A \times B = \{ac, ad, bc, bd\}$$

1.1.2 Формальное определение языка.

Понятие языка

- **Языком L** над алфавитом V называют некоторое счетное подмножество цепочек конечной длины из множества всех цепочек над алфавитом V . $L(V) \subseteq V^*$
 - ✓ множество цепочек языка не обязано быть конечным
 - ✓ хотя каждая цепочка в языке обязана быть конечной длины, эта длина формально ничем не ограничена
- **Предложением языка** называется цепочка символов, принадлежащая этому языку.

1.1.2 Формальное определение языка.

Понятие языка

- Язык L над алфавитом V включает в себя язык L' над алфавитом V ($L(V) \leq L'(V)$), если справедливо, что любая цепочка α принадлежащая L , принадлежит и L' . $\alpha \in L(V)$ и $\alpha \in L'(V)$
- Два языка $L(V)$ и $L'(V)$ **равны** или совпадают если справедливо $L(V) \leq L'(V)$ и $L'(V) \leq L(V)$.
- Два языка $L(V)$ и $L'(V)$ **почти эквиваленты**, если они отличаются на пустую цепочку $L(V) =_{\sim} L'(V)$.

$$L(V) \cup \{\varepsilon\} = L'(V) \cup \{\varepsilon\} .$$

1.1.3 Способы задания языка

- перечисление всех допустимых цепочек языка
- с помощью указания способа порождения цепочек языка (задать грамматику языка, используемую для порождения языка)

$$L(G1) = \{0^n 1^n \mid n > 0\}$$

01

0011

000111

$$L(G1) = \{0^n 1^m \mid n > 0, m > 0\}$$

01

0000011

00111

- определение метода распознавания цепочек языка

1.1.4 Синтаксис и семантика

- **Лексема** – это языковая конструкция, которая состоит из элементов алфавита языка и не содержит других конструкций.
- **Синтакис** – набор правил, определяющих допустимые конструкции языка. Синтаксис определяет форму языка.
- **Семантика** – это раздел языка, определяющий значения предложений языка (определяющий содержание, смысл языка).

ГЛАВА 1. Основы теории формальных языков и грамматик

1.2 ОПРЕДЕЛЕНИЕ ГРАММАТИКИ

1.2.1 Понятие о грамматике языка

- **Грамматика** – описание способов построения предложений некоторого языка.
- Грамматика — один из основных подходов к описанию бесконечного формального языка конечными средствами.
- **Правило (продукция)** – упорядоченная пара цепочек $(\alpha \beta)$, которое записывается $\alpha \rightarrow \beta$ (α порождает β).
- $L(G)$ – язык L , заданный грамматикой G .

1.2.1 Понятие о грамматике языка

- Грамматиками $G1$ и $G2$ называются **эквивалентными**, если $L(G1) = L(G2)$.

$$G1 = (\{0,1\}, \{A,S\}, P1, S)$$

$$P1: \quad S \rightarrow 0A1$$

$$0A \rightarrow 00A1$$

$$A \rightarrow \varepsilon$$

$$G2 = (\{0,1\}, \{S\}, P2, S)$$

$$P2: \quad S \rightarrow 0S1 \mid 01$$

$$L(G1) = L(G2) = \{0^n 1^n \mid n > 0\}.$$

- Грамматиками $G1$ и $G2$ называются **почти эквивалентными**, если $L(G1) \cup \{\varepsilon\} = L(G2) \cup \{\varepsilon\}$.

$$G1 = (\{0,1\}, \{A,S\}, P1, S)$$

$$P1: \quad S \rightarrow 0A1$$

$$0A \rightarrow 00A1$$

$$A \rightarrow \varepsilon$$

$$L(G1) = \{0^n 1^n \mid n > 0\}$$

$$G2 = (\{0,1\}, \{S\}, P2, S)$$

$$P2: \quad S \rightarrow 0S1 \mid \varepsilon$$

$$L(G2) = \{0^n 1^n \mid n \geq 0\}$$

1.2.2 Формальное определение грамматики

По определению Хомского формальная грамматика представляет собой четвёрку:

$$G = \{V_T, V_N, P, S\}$$

- V_T , T - множество терминальных символов языка,
- V_N , N – множество нетерминальных символов (или понятий языка или синтаксических единиц)

$$V = V_N \cup V_T$$

$$V_N \cap V_T = \emptyset$$

- P – множество правил подстановки (продукций), имеющий вид $\alpha \rightarrow \beta$, $\alpha \in V^+$, $\beta \in V^*$.

Знак \rightarrow означает "непосредственно порождает" или "есть по определению".

- S – аксиома грамматики или начальный символ грамматики. $S \in V_N$.

1.1.2 Формальное определение грамматики

Грамматика, определяющая целое число без знака:

$$G = \{VT, VN, P, S\}$$

$$VN = \{A, B\}$$

$$VT = \{0, 1, 2, 3, 4, 5, 6, 7, 8, 9\}$$

$$P = \{A \rightarrow BA, A \rightarrow B, B \rightarrow 0, B \rightarrow 1, B \rightarrow 9\}$$

$$S = \{A\}$$

А - целое число без знака, В - любая цифра.

ГЛАВА 1. Основы теории формальных языков и грамматик

1.3 СПОСОБЫ ЗАПИСИ СИНТАКСИСА **ЯЗЫКА**

*Метаязык - язык,
предназначенный для
описания другого языка*

1.3.1 Метаязык Хомского

- -> символ отделяет левую часть правила от правой (читается как "порождает" и "это есть");
- нетерминалы обозначаются буквой A с индексом, указывающим на его номер;
- терминалы - это символы, используемые в описываемом языке;
- Каждое правило определяет порождение одной новой цепочки, причём один и тот же нетерминал может встречаться в нескольких правилах слева.

1.3.1 Метаязык Хомского

- | | | | |
|-------------------------|-------------------------|-------------------------|-------------------------------|
| 1. $A_1 \rightarrow A$ | 17. $A_1 \rightarrow Q$ | 33. $A_1 \rightarrow g$ | 49. $A_1 \rightarrow w$ |
| 2. $A_1 \rightarrow B$ | 18. $A_1 \rightarrow R$ | 34. $A_1 \rightarrow h$ | 50. $A_1 \rightarrow x$ |
| 3. $A_1 \rightarrow C$ | 19. $A_1 \rightarrow S$ | 35. $A_1 \rightarrow i$ | 51. $A_1 \rightarrow y$ |
| 4. $A_1 \rightarrow D$ | 20. $A_1 \rightarrow T$ | 36. $A_1 \rightarrow j$ | 52. $A_1 \rightarrow z$ |
| 5. $A_1 \rightarrow E$ | 21. $A_1 \rightarrow U$ | 37. $A_1 \rightarrow k$ | 53. $A_2 \rightarrow 0$ |
| 6. $A_1 \rightarrow F$ | 22. $A_1 \rightarrow V$ | 38. $A_1 \rightarrow l$ | 54. $A_2 \rightarrow 1$ |
| 7. $A_1 \rightarrow G$ | 23. $A_1 \rightarrow W$ | 39. $A_1 \rightarrow m$ | 55. $A_2 \rightarrow 2$ |
| 8. $A_1 \rightarrow H$ | 24. $A_1 \rightarrow X$ | 40. $A_1 \rightarrow n$ | 56. $A_2 \rightarrow 3$ |
| 9. $A_1 \rightarrow I$ | 25. $A_1 \rightarrow Y$ | 41. $A_1 \rightarrow o$ | 57. $A_2 \rightarrow 4$ |
| 10. $A_1 \rightarrow J$ | 26. $A_1 \rightarrow Z$ | 42. $A_1 \rightarrow p$ | 58. $A_2 \rightarrow 5$ |
| 11. $A_1 \rightarrow K$ | 27. $A_1 \rightarrow a$ | 43. $A_1 \rightarrow q$ | 59. $A_2 \rightarrow 6$ |
| 12. $A_1 \rightarrow L$ | 28. $A_1 \rightarrow b$ | 44. $A_1 \rightarrow r$ | 60. $A_2 \rightarrow 7$ |
| 13. $A_1 \rightarrow M$ | 29. $A_1 \rightarrow c$ | 45. $A_1 \rightarrow s$ | 61. $A_2 \rightarrow 8$ |
| 14. $A_1 \rightarrow N$ | 30. $A_1 \rightarrow d$ | 46. $A_1 \rightarrow t$ | 62. $A_2 \rightarrow 9$ |
| 15. $A_1 \rightarrow O$ | 31. $A_1 \rightarrow e$ | 47. $A_1 \rightarrow u$ | 63. $A_3 \rightarrow A_1$ |
| 16. $A_1 \rightarrow P$ | 32. $A_1 \rightarrow f$ | 48. $A_1 \rightarrow v$ | 64. $A_3 \rightarrow A_3 A_1$ |
| | | | 65. $A_3 \rightarrow A_3 A_2$ |

Описание идентификатора на метаязыке Хомского

1.3.2 Бэкуса-Наура формы (БНФ)

- символ "::<=" отделяет левую часть правила от правой;
- нетерминалы обозначаются произвольной символьной строкой, заключённой в угловые скобки "<" и ">";
- терминалы – это символы, используемые в описываемом языке;
- каждое правило определяет порождение нескольких альтернативных цепочек, отделяемых друг от друга символом вертикальной черты "|".

1.3.2 Бэкуса-Наура формы (БНФ)

1. $\langle \text{буква} \rangle ::= A \mid B \mid C \dots \mid Z \mid a \mid b \mid c \mid \dots \mid z$

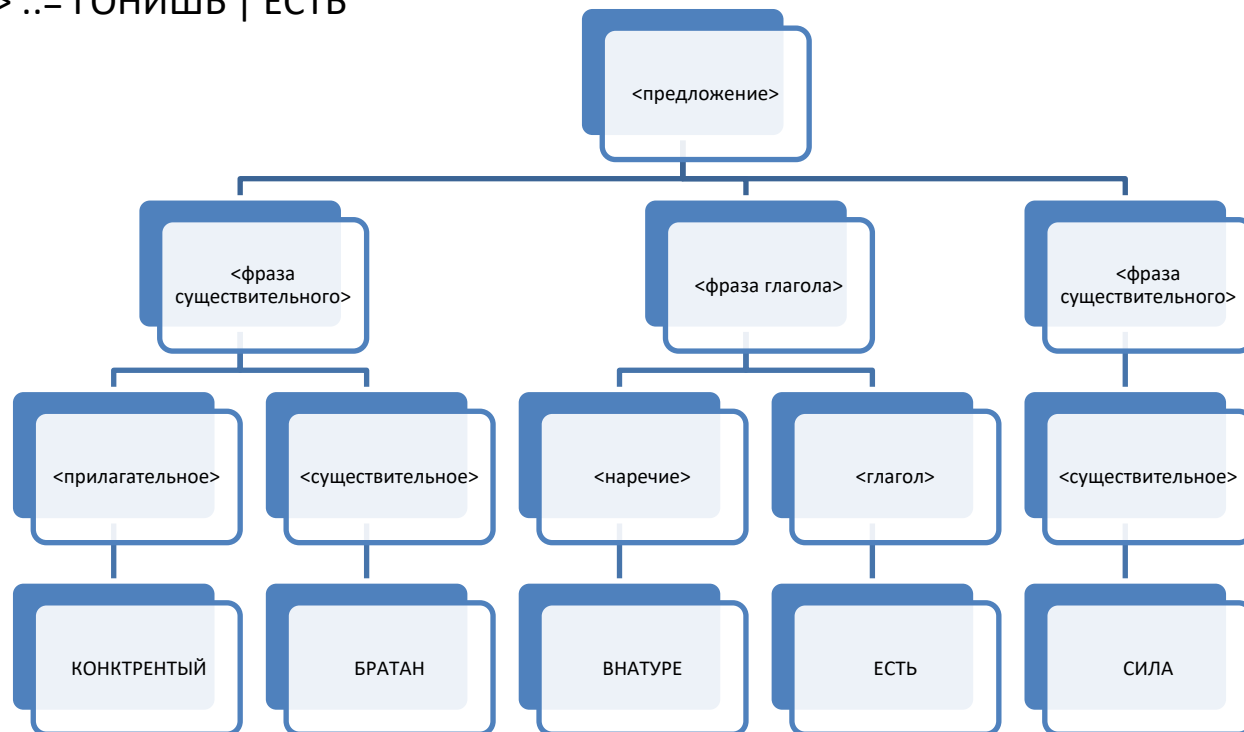
1. $\langle \text{цифра} \rangle ::= 0 \mid 1 \mid 2 \dots \mid 9$

3. $\langle \text{идентификатор} \rangle ::= \langle \text{буква} \rangle \mid$
 $\quad \langle \text{идентификатор} \rangle \langle \text{буква} \rangle \mid$
 $\quad \langle \text{идентификатор} \rangle \langle \text{цифра} \rangle$

Пример описания идентификатора с использованием БНФ

1.3.2 Бэкуса-Наура формы (БНФ)

- $\langle \text{предложение} \rangle ::= \langle \text{фраза существительного} \rangle \langle \text{фраза глагола} \rangle \langle \text{фраза существительного} \rangle$
- $\langle \text{фраза существительного} \rangle ::= \langle \text{прилагательное} \rangle \langle \text{существительное} \rangle \mid \langle \text{существительное} \rangle$
- $\langle \text{прилагательное} \rangle ::= \text{БЛАТНОЙ} \mid \text{КОНКРЕТНЫЙ}$
- $\langle \text{существительное} \rangle ::= \text{ПАЦАН} \mid \text{БРАТАН} \mid \text{СИЛА}$
- $\langle \text{фраза глагола} \rangle ::= \langle \text{наречие} \rangle \langle \text{глагол} \rangle \mid \langle \text{глагол} \rangle$
- $\langle \text{наречие} \rangle ::= \text{ЧИСТО} \mid \text{КОНКРЕТНО} \mid \text{ТИПА} \mid \text{ВНАТУРЕ}$
- $\langle \text{глагол} \rangle ::= \text{ГОНИШЬ} \mid \text{ЕСТЬ}$



Упрощенная грамматика русского языка в терминах БНФ

1.3.3 РБНФ (расширенная)

- [] – синтаксическая конструкция может отсутствовать;
- { } – повторение синтаксической конструкции (возможно 0 раз)
- () – для ограничения альтернативных конструкций
- { \ \ } – для обозначения повторения один и более раз.

1.3.3 РБНФ

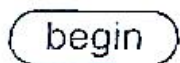
- $\langle \text{буква} \rangle ::= A \mid B \mid C \dots \mid Z \mid a \mid b \mid c \mid \dots \mid z$
- $\langle \text{цифра} \rangle ::= 0 \mid 1 \mid 2 \dots \mid 9$
- $\langle \text{идентификатор} \rangle ::= \langle \text{буква} \rangle \{ \langle \text{буква} \rangle \mid \langle \text{цифра} \rangle \}$

Пример описания идентификатора с использованием РБНФ

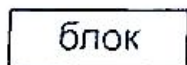
1.3.4 Диаграмма Вирта



терминальный символ, принадлежащий алфавиту языка



постоянная группа терминальных символов, определяющая название лексемы, ключевое слово и т.д.



нетерминальный символ определяющий название правила

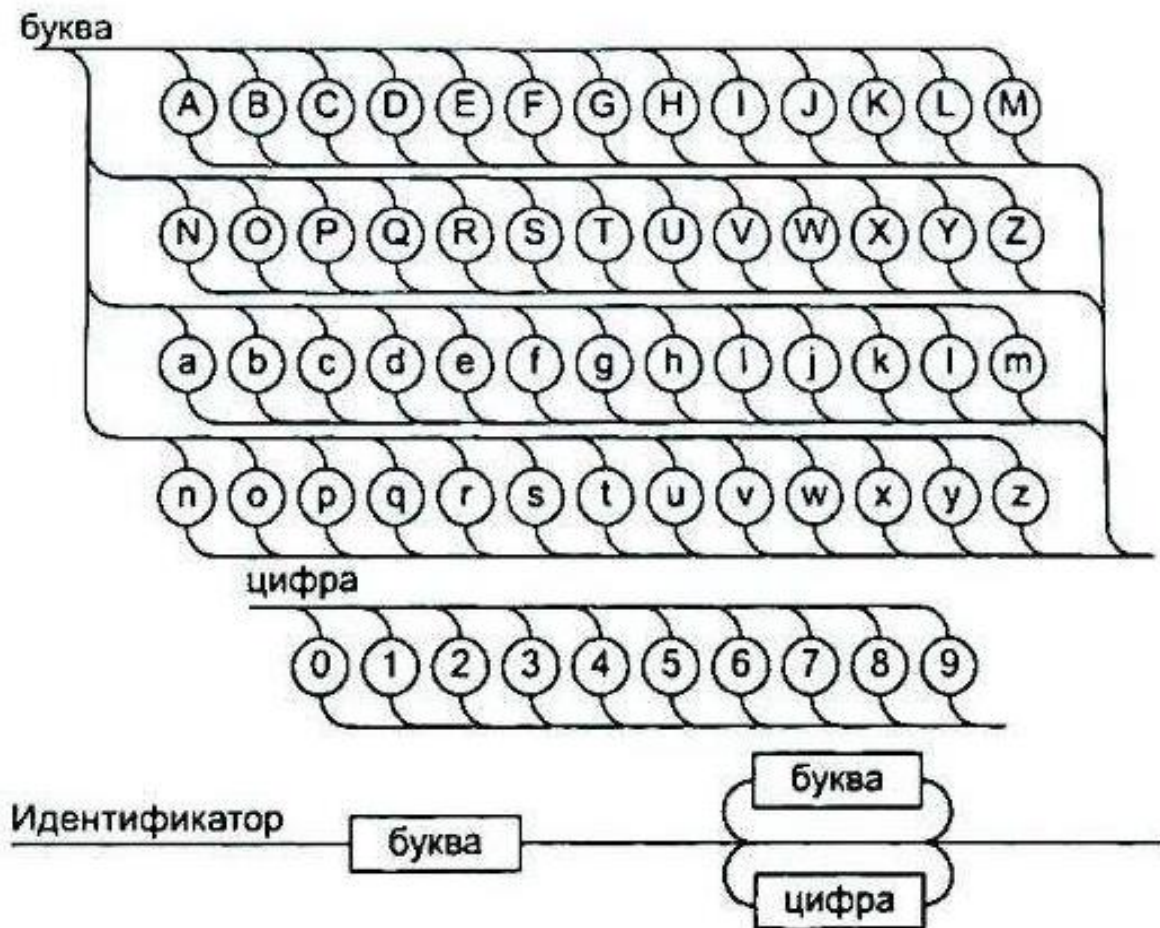


соединительные линии, обеспечивающие связь между терминальными и нетерминальными символами в правилах, заданных диаграммами Вирта



входная дуга с именем правила, определяющая его начало

1.3.4 Диаграмма Вирта



ГЛАВА 1. Основы теории формальных языков и грамматик

1.4 КЛАССИФИКАЦИЯ ЯЗЫКОВ И ГРАММАТИК

1.4.1 Классификация грамматик

Тип	Название	Ограничение на правила
0	Грамматика с фразовой структурой (грамматика без ограничений)	$\alpha \rightarrow \beta,$ где $\alpha \in V^+, \beta \in V^*$
1	Контекстно-зависимые неукорачивающие	$\xi_1 A \xi_2 \rightarrow \xi_1 \gamma \xi_2,$ где $A \in VN; \gamma \in V^+; \xi_1, \xi_2 \in V^*.$ $\alpha \rightarrow \beta,$ где $\alpha, \beta \in V^+$ и $ \alpha \leq \beta $

1.4.1 Классификация грамматик

Тип	Название	Ограничение на правила
2	Контекстно-свободные: неукорачивающие укорачивающие	$A \rightarrow \beta$, где $A \in VN$, $\beta \in V^+$ $A \rightarrow \beta$, где $A \in VN$, $\beta \in V^*$
3	Регулярные: леволинейные праволинейные Подкласс регулярных – автоматные: леволинейные праволинейные	$A \rightarrow B\gamma \mid \gamma$, где $A, B \in VN$, $\gamma \in VT^*$ $A \rightarrow \gamma B \mid \gamma$, где $A, B \in VN$, $\gamma \in VT^*$ $A \rightarrow Bt \mid t$, где $A, B \in VN$, $t \in VT$ $A \rightarrow tB \mid t$, где $A, B \in VN$, $t \in VT$

1.4.1 Классификация грамматик

- Эта иерархия грамматик – включающая.
- Грамматика 2 включает 3, но не наоборот.
- Любая грамматика относится к типу 0.
- Существуют такие УКС грамматики, которые не относятся к КЗ и неукорачивающим, а относятся к типу без ограничений.
- Сложность грамматики обратно пропорциональна тому максимально возможному номеру типа к которому может быть отнесена грамматика.

1.4.2 Классификация языков

Языки классифицируются в соответствие с типами грамматик с помощью которых они заданы. Поскольку один и тот же язык в общем случае может быть задан сколь угодно количеством грамматик, которые могут относиться к разным типам, то для классификации языка всегда выбирается грамматика с максимальным классификационным типом.

1.4.2 Классификация языков

Грамматика 0 $G1 = (\{0,1\}, \{A,S\}, P1, S)$ и

$P1: \quad S \rightarrow 0A1$

$0A \rightarrow 00A1$

$A \rightarrow \varepsilon$

КС-грамматика $G2 = (\{0,1\}, \{S\}, P2, S)$, где

$P2: \quad S \rightarrow 0S1 \mid 01$

описывают один и тот же язык

$$L = L(G1) = L(G2) = \{ 0^n 1^n \mid n > 0 \}$$

1.4.2 Классификация языков

- Сложность языка убывает с возрастанием классификационного типа языка.
- Тип 0. Язык с фразовой структурой (естественные языки).
- Тип 1. Язык контекстно-зависимый.

В общем случае время на распознавание предложения этого языка экспоненциально зависит от длины входящей цепочки.
- Тип 2. Контекстно-свободный язык.

Время распознавания предложений КС-языка полиномиально зависит от длины входящей цепочки.
- Тип 3. Регулярные.

Время распознавания предложений языка линейно зависит от длины входящей цепочки.

1.4.3 Примеры классификации языков и грамматик

- Язык типа 2: $L(G3) = \{(ac)^n (cb)^n \mid n > 0\}$

G3: $S \rightarrow aQb \mid accb$

$Q \rightarrow cSc$

- Язык типа 3: $L(G4) = \{\omega \perp \mid \omega \in \{a,b\}^+, \text{ где нет двух рядом стоящих } a\}$

G4: $S \rightarrow A\perp \mid B\perp$

$A \rightarrow a \mid Ba$

$B \rightarrow b \mid Bb \mid Ab$

1.4.3 Примеры классификации языков и грамматик

Язык типа 0:

$$L(G1) = \{a^2 b^{n^2-1} \mid n \geq 1\}$$

G1: $S \rightarrow aaCFD$

$F \rightarrow AFB \mid AB$

$AB \rightarrow bBA$

$Ab \rightarrow bA$

$AD \rightarrow D$

$Cb \rightarrow bC$

$CB \rightarrow C$

$bCD \rightarrow \varepsilon$

Язык типа 1:

$$L(G2) = \{a^n b^n c^n, n \geq 1\}$$

G2: $S \rightarrow aSBC \mid abC$

$CB \rightarrow BC$

$bB \rightarrow bb$

$bC \rightarrow bc$

$cC \rightarrow cc$

ГЛАВА 1. Основы теории формальных языков и грамматик

1.5 ЦЕПОЧКИ ВЫВОДА. **СЕНТЕНЦИАЛЬНАЯ ФОРМА**

1.5.1 Вывод. Цепочка вывода.

- **Выводом** называется процесс порождения предложений языка на основе правил, определяющих язык.
- Цепочка $\beta \in (VT \cup VN)^*$ **непосредственно выводима** из цепочки $\alpha \in (VT \cup VN)^+$ в грамматике $G = (VT, VN, P, S)$ (обозначим $\alpha \Rightarrow \beta$), если $\alpha = \xi_1 \gamma \xi_2$, $\beta = \xi_1 \delta \xi_2$, где $\xi_1, \xi_2, \delta \in (VT \cup VN)^*$, $\gamma \in (VT \cup VN)^+$ и правило вывода $\gamma \rightarrow \delta$ содержится в P .

1.5.1 Вывод. Цепочка вывода.

- ❑ Цепочка $\beta \in V^*$ выводима из цепочки $\alpha \in V^+$ в грамматике $G = (V_T, V_N, P, S)$ (обозначим $\alpha \Rightarrow^* \beta$), если:
 - β непосредственно выводима из α ($\alpha \Rightarrow \beta$)
 - существует такая цепочка γ , что β непосредственно выводима из γ ($\gamma \Rightarrow \beta$), а γ выводима из α ($\alpha \Rightarrow^* \gamma$)
- ❑ **β выводима из α** если β непосредственно выводима из α или если можно построить цепочку непосредственных выводов $\alpha \Rightarrow \gamma_0 \Rightarrow \gamma_1 \Rightarrow \dots \Rightarrow \gamma_n \Rightarrow \beta$.
- ❑ Такая последовательность непосредственно выводимых цепочек называется **цепочкой вывода**.
- ❑ Каждый переход от одной цепочки к другой называется **шагом вывода**.

1.5.1 Вывод. Цепочка вывода.

- Если цепочка вывода от α к β содержит одну и более промежуточных цепочек, то такая цепочка обозначается $\alpha \Rightarrow^+ \beta$ (**β нетривиально выводима из α**).
- Если количество шагов вывода известно, то его можно указать непосредственно над знаком вывода.

Если $\alpha \Rightarrow \beta$, то один шаг вывода.

Если $\alpha \Rightarrow^4 \beta$, то β выводима из α за 4 шага.

Если $\alpha \Rightarrow^0 \beta$, то $\alpha = \beta$.

1.5.1 Вывод. Цепочка вывода.

$$G = (\{A, S\}, \{0, 1\}, P, S)$$

$$P: S \rightarrow 0A1$$

$$0A \rightarrow 00A1$$

$$A \rightarrow \varepsilon$$

Приведем вывод для цепочки $0011 \in L(G)$:

- $S \Rightarrow 0A1 \Rightarrow 00A11 \Rightarrow 0011$;
- $S \Rightarrow^+ 0A1$; $S \Rightarrow^+ 00A11$; $S \Rightarrow^+ 0011$;
- $S \Rightarrow^* S$; $S \Rightarrow^m 0A1$; $S \Rightarrow^* 00A11$; $S \Rightarrow^3 0011$;

1.5.2 Сентенциальная форма грамматики. Основа

- Вывод называется законченным, если на основе цепочки β , полученной в результате вывода нельзя сделать ни одного шага вывода, т.е. цепочка β состоит из терминальных символов.
- Цепочка β , полученная в результате законченного вывода называется **конечной цепочкой** вывода.
- Цепочка $\alpha \in (VT \cup VN)^*$, для которой $S \Rightarrow^* \alpha$ (если α выводима из начального символа грамматики), называется **сентенциальной формой** в грамматике $G = (VT, VN, P, S)$.
- Если $\alpha \in VT^*$, то α называется **конечной сентенциальной формой** или предложением языка.

1.5.2 Сентенциальная форма грамматики. Основа

- Пусть $G=(VN, VT, P, S)$ грамматика и цепочка $w = \gamma_1 \beta \gamma_2$ сентенциальная форма $\gamma_1, \gamma_2 \in V^*$, $\beta \in V^+$, тогда β называют **фразой сентенциальной формы** w для нетерминала B , если существуют выводы $S \Rightarrow^* \gamma_1 B \gamma_2$, $B \Rightarrow^+ \beta$.
- β называется **простой фразой**, если существуют выводы $S \Rightarrow^* \gamma_1 B \gamma_2$, $B \Rightarrow \beta$.
- Основой всякой сентенциальной формы называется самая левая простая фраза.
если $\gamma_1 = \varepsilon$, то B - голова.
если $\gamma_2 = \varepsilon$, то B - хвост.
- Язык L заданный грамматикой G - это множество всех конечных сентенциальных форм грамматики G .

1.5.3 Левосторонний и правосторонний вывод

- Вывод цепочки $\beta \in (VT)^*$ из $S \in VN$ в КС-грамматике $G = (VT, VN, P, S)$, называется левым (левосторонним), если в этом выводе каждая очередная сентенциальная форма получается из предыдущей заменой самого левого нетерминала.
- Вывод цепочки $\beta \in (VT)^*$ из $S \in VN$ в КС-грамматике $G = (VT, VN, P, S)$, называется правым (правосторонним), если в этом выводе каждая очередная сентенциальная форма получается из предыдущей заменой самого правого нетерминала.

1.5.3 Левосторонний и правосторонний вывод

Например, для цепочки $a+b+a$ в грамматике

$G = (\{a,b,+\}, \{S,T\}, \{S \rightarrow T \mid T+S; T \rightarrow a \mid b\}, S)$

можно построить выводы:

(1) $S \Rightarrow T+S \Rightarrow T+T+S \Rightarrow T+T+T \Rightarrow a+T+T \Rightarrow a+b+T \Rightarrow a+b+a$

(2) $S \Rightarrow T+S \Rightarrow a+S \Rightarrow a+T+S \Rightarrow a+b+S \Rightarrow a+b+T \Rightarrow a+b+a$

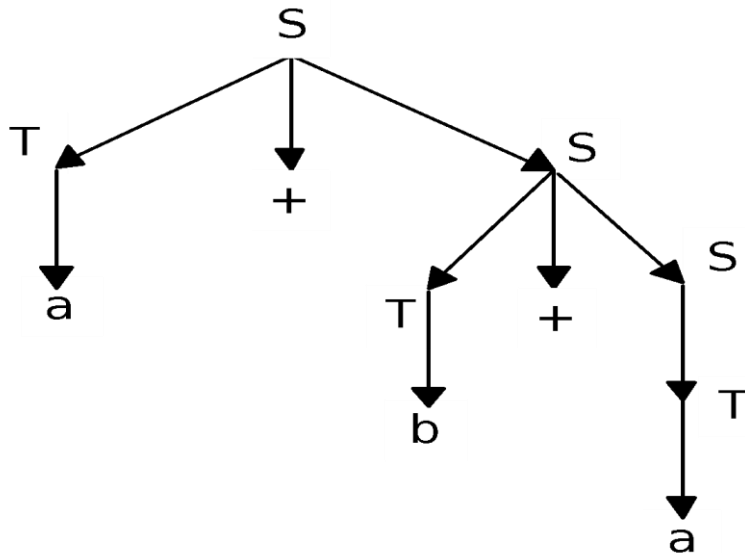
(3) $S \Rightarrow T+S \Rightarrow T+T+S \Rightarrow T+T+T \Rightarrow T+T+a \Rightarrow T+b+a \Rightarrow a+b+a$

Для грамматик типов 2,3 для любой сентенциальной формы можно построить левый и правый вывод.

Для грамматик типов 0,1 – не всегда.

1.5.4 Дерево вывода

- Цепочку вывода можно представить графически в форме дерева (графа).
- Корнем дерева является вершина, обозначенная начальным символом грамматики.
- В узлах дерева находятся нетерминальные символы, в листьях – терминалы.
- Каждая связь соответствует одному шагу вывода.



Процесс распознавания цепочки символов – можно ли построить для данной цепочки дерево назовем синтаксическим анализом, а само дерево – синтаксическим.