(1) Примеры грамматик и языков

$$
\begin{aligned}
S &\to ABCS \mid ABc \\
BA &\to AB \\
CA &\to AC \\
CB &\to BC \\
Cc &\to cc \\
Bc &\to bc \\
Bb &\to bb \\
Ab &\to ab \\
Aa &\to aa
\end{aligned}
$$

Тип 1. Неукорачивающая, но не КЗ

Язык: $\{a^n b^n c^n \mid n > 0\}$

Примеры грамматик и языков

(2)

$$S \to aSb \mid ab$$

Язык: $\{a^n b^n \mid n > 0\}$

(3)

$$S \to aS \mid a$$

Язык: $\{a^n \mid n > 0\}$

Иерархия классов Хомского

Задача распознавания

Даны грамматика G и цепочка x

$x \in L(G)$?

Для грамматик типа 1 (а также типов 2 и 3) по классификации Хомского задача распознавания разрешима, т.е. существует общий алгоритм, отвечающий на вопрос: $x \in L(G)$?

Контекстно-свободные грамматики и языки

КС-грамматики позволяют выразить такие свойства языков программирования, как скобочные структуры, последовательность описаний и операторов и др. Но не могут задавать контекстно-зависимые свойства, например, соответствие числа формальных и фактических параметров при вызове функции. Для КС-грамматик существуют эффективные алгоритмы анализа, поэтому они применяются в трансляции, контекстные условия проверяются на

этапе семантического анализа

Левый (левосторонний) вывод цепочки $\beta \in T^*$ из $S \in N$ в КС-грамматике $G = (T, N, P, S)$:

в этом выводе каждая очередная сентенциальная форма получается из предыдущей заменой самого левого нетерминала.

Правый (правосторонний) вывод цепочки $\beta \in T^*$ из $S \in N$ в КС-грамматике $G = (T, N, P, S)$:

в этом выводе каждая очередная сентенциальная форма получается из предыдущей заменой самого правого нетерминала.

Рассмотрим пример грамматики:

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

можно построить выводы для цепочки a+b+a :

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

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

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

Здесь (2) - левосторонний вывод, (3) - правосторонний, а (1) не является ни левосторонним, ни правосторонним

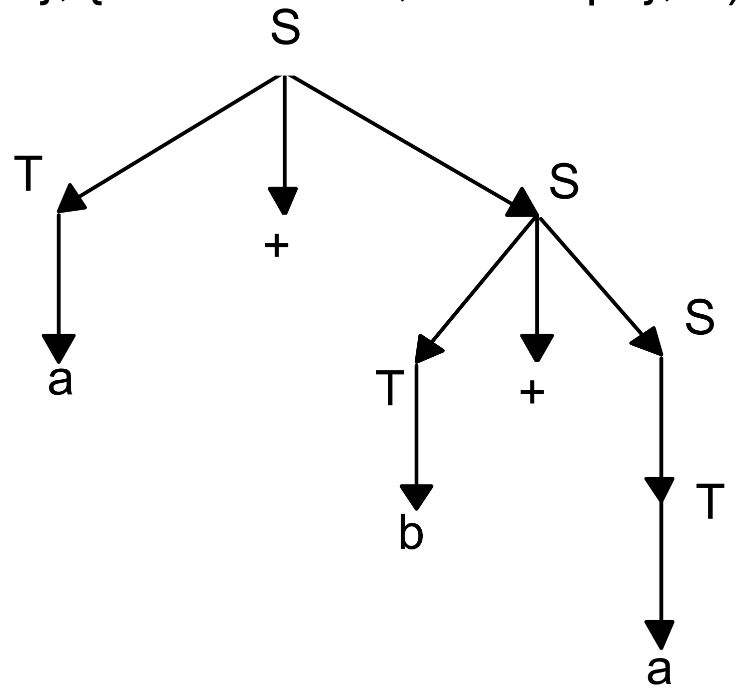
***Определение*: упорядоченное ориентированное дерево называется **деревом вывода** (или **деревом разбора**) в КС-грамматике G = (T, N, P, S), если выполнены следующие условия:

(1) каждая вершина дерева помечена символом из множества $N \cup T \cup \{\varepsilon\}$, при этом корень дерева помечен символом S; листья - символами из $T \cup \{\varepsilon\}$;

(2) если вершина дерева помечена символом A , а ее непосредственные потомки - символами a_1, a_2, ... , a_n, где каждое $a_i \in T \cup N$, то $A \to a_1 a_2 ... a_n$ - правило вывода в этой грамматике;

(3) если вершина дерева помечена символом A , а ее единственный непосредственный потомок помечен символом ε, то $A \to \varepsilon$ — правило вывода в этой грамматике.

Пример дерева вывода для цепочки a+b+a в грамматике G = ({a,b,+}, {S,T}, {S $\rightarrow$ T | T+S; T $\rightarrow$ a | b}, S) :



(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

КС-грамматика G называется **неоднозначной**, если существует хотя бы одна цепочка $\alpha \in L(G)$, для которой может быть построено два или более различных деревьев вывода.

Это утверждение эквивалентно тому, что цепочка α имеет два или более разных левосторонних (или правосторонних) выводов.

В противном случае грамматика называется **однозначной.**

Утв. Проблема определения, является ли заданная КС-грамматика однозначной, является **алгоритмически неразрешимой**.

Язык, порождаемый грамматикой, называется **неоднозначным**, если он не может быть порожден никакой однозначной грамматикой.

Утв. Проблема определения, порождает ли данная КС-грамматика однозначный язык (т.е. существует ли эквивалентная ей однозначная грамматика), является **алгоритмически неразрешимой.**

- Пример неоднозначного языка:

$L= \{a^n\ b^n\ c^m \mid n>0,\ m>0\} \cup \{a^n\ b^m\ c^m \mid n>0, m>0\}$