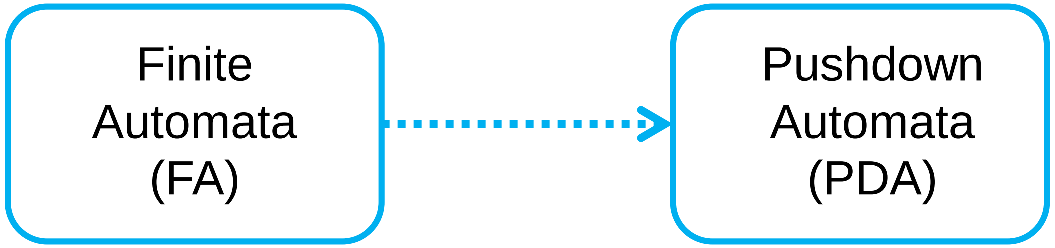
بسم الله الرحمن الرحیم

نظریه زبان‌ها و ماشین‌ها

جلسه ۱۳

مجتبی خلیلی
دانشکده برق و کامپیوتر
دانشگاه صنعتی اصفهان

سه مدل



- اتوماتای پشته‌ای
  - حافظه نامحدود اما دسترسی خاص (stack)
  - زبان‌ها و گرامرهای مستقل از متن
  - زبان‌های برنامه نویسی، کامپایلر

مثال

- زبان زیر منظم است یا نامنظم؟

$$\{0^n 1^n : n \geq 0\} = \{\varepsilon, 01, 0011, 000111, \ldots\}$$

زبان‌های مستقل از متن

- زبان زیر را در نظر بگیرید:

$$L = \{w \in \{(,)\}^* \mid w \text{ is in a balanced form}\}$$

$$L \ni \epsilon,\ (),\ (()),\ ()(),\ (()()),\ (()(())),\ ((())),\ \cdots$$
$$L \not\ni (,\ ),\ )(,\ ((),\ ()),\ (())),\ (()((),\ \cdots$$

- این زبان منظم نیست (خودتان میتوانید با PL بررسی کنید).

- این زبان را چگونه توصیف کنیم؟

گرامر

IUT-ECE

- گرامر و PDA

گرامر

- مثال: حالتی ساده از گرامر زبان انگلیسی

$$
\begin{aligned}
\langle \text{SENTENCE} \rangle &\to \langle \text{NOUN-PHRASE} \rangle \langle \text{VERB} \rangle \\
\langle \text{NOUN-PHRASE} \rangle &\to \langle \text{CMPLX-NOUN} \rangle \\
\langle \text{CMPLX-NOUN} \rangle &\to \langle \text{ARTICLE} \rangle \langle \text{NOUN} \rangle \\
\langle \text{ARTICLE} \rangle &\to \texttt{a} \mid \texttt{the} \\
\langle \text{NOUN} \rangle &\to \texttt{boy} \mid \texttt{girl} \mid \texttt{flower} \\
\langle \text{VERB} \rangle &\to \texttt{touches} \mid \texttt{likes} \mid \texttt{sees}
\end{aligned}
$$

گرامر

- مثال: اشتقاق یک جمله

$$
\begin{aligned}
\langle \text{SENTENCE} \rangle &\Rightarrow \langle \text{NOUN-PHRASE} \rangle \langle \text{VERB} \rangle \\
&\Rightarrow \langle \text{CMPLX-NOUN} \rangle \langle \text{VERB} \rangle \\
&\Rightarrow \langle \text{ARTICLE} \rangle \langle \text{NOUN} \rangle \langle \text{VERB} \rangle \\
&\Rightarrow \texttt{a} \ \langle \text{NOUN} \rangle \langle \text{VERB} \rangle \\
&\Rightarrow \texttt{a boy} \ \langle \text{VERB} \rangle \\
&\Rightarrow \texttt{a boy sees}
\end{aligned}
$$

گرامر

- مثال: اشتقاق یک جمله

$$\begin{aligned}
\langle\text{SENTENCE}\rangle &\Rightarrow \langle\text{NOUN-PHRASE}\rangle\langle\text{VERB}\rangle \\
&\Rightarrow \langle\text{CMPLX-NOUN}\rangle\langle\text{VERB}\rangle \\
&\Rightarrow \langle\text{ARTICLE}\rangle\langle\text{NOUN}\rangle\langle\text{VERB}\rangle \\
&\Rightarrow \texttt{a}\ \langle\text{NOUN}\rangle\langle\text{VERB}\rangle \\
&\Rightarrow \texttt{a boy}\ \langle\text{VERB}\rangle \\
&\Rightarrow \texttt{a boy sees}
\end{aligned}$$

زبان

گرامر

- یک گرامر مجموعه‌ای از قواعد برای ساخت رشته‌ها/ساخت یک زبان است. در واقع، گرامرها یک شیوه ممکن برای تولید یا مشخص کردن زبان‌ها هستند.
- یک گرامر شامل:
  - یک مجموعه از متغیرهاست (شامل متغیر آغازین)
  - یک مجموعه از ترمینال‌هاست (از الفبا)
  - یک لیست از قواعد
- مثال:

$$S \to 0S1$$
$$S \to \epsilon$$

$\longrightarrow$ 0^n1^n

گرامر

- مثال:

$$S \rightarrow 0S1$$
$$S \rightarrow \epsilon$$

$$\longrightarrow \quad 0^n1^n$$

- گوییم رشته w (فقط شامل ترمینالها) توسط گرامر مورد نظر تولید شده است اگر با شروع از متغیر آغازین (S=start) و اعمال قواعد بتوان w را بدست آورد.

- مثلا رشته زیر توسط گرامر بالا تولید شده است. به دنباله زیر که نشان دهنده روند تولید این رشته است اشتقاق (derivation) گویند.

$$S \Rightarrow 0S1 \Rightarrow 00S11 \Rightarrow 0011$$

گرامر

- نمادها در اشتقاق:

یک گام	$S \Rightarrow w$
صفر گام یا بیشتر	$S \Rightarrow^{*} w$
یک گام یا بیشتر	$S \Rightarrow^{+} w$

مثال

- رشته aabbcc را از گرامر زیر اشتقاق کنید:

$$V = \{\mathsf{S}, \mathsf{X}, \mathsf{Y}\}$$
$$\Sigma = \{\mathtt{a}, \mathtt{b}, \mathtt{c}\}.$$

$$\mathsf{S} \to \varepsilon \qquad \mathsf{X} \to \mathtt{a}\mathsf{X}\mathsf{Y}\mathtt{c} \qquad \mathtt{c}\mathsf{Y} \to \mathsf{Y}\mathtt{c}$$
$$\mathsf{S} \to \mathtt{abc} \qquad \mathsf{X} \to \mathtt{abc} \qquad \mathtt{b}\mathsf{Y} \to \mathtt{bb}$$
$$\mathsf{S} \to \mathsf{X}$$

تعریف فرمال گرامر

IUT-ECE

DEFINITION 1.1

A grammar G is defined as a quadruple

$$G = (V, T, S, P),$$

where V is a finite set of objects called **variables**,
T is a finite set of objects called **terminal symbols**,
$S \in V$ is a special symbol called the **start** variable,
P is a finite set of **productions**.

It will be assumed without further mention that the sets V and T are nonempty and disjoint.

$$\boxed{P \;:\; (V \cup T)^{+} \longrightarrow (V \cup T)^{*}.}$$

زبان تولید شده با گرامر

IUT-ECE

DEFINITION 1.2

Let $G = (V, T, S, P)$ be a grammar. Then the set

$$L(G) = \left\{ w \in T^* : S \stackrel{*}{\Rightarrow} w \right\}$$

is the language generated by G.

مثال

- گرامر زیر چه زبانی را تولید می‌کند؟

$$S \rightarrow 0S1 \mid A$$
$$A \rightarrow \#$$

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

مثال

EXAMPLE 1.12

Find a grammar that generates

$$L = \{a^n b^{n+1} : n \geq 0\}.$$

The idea behind the previous example can be extended to this case. All we need to do is generate an extra b. This can be done with a production $S \rightarrow Ab$, with other productions chosen so that A can derive the language in the previous example. Reasoning in this fashion, we get the grammar $G = (\{S, A\}, \{a, b\}, S, P)$, with productions

$$\begin{aligned} S &\rightarrow Ab, \\ A &\rightarrow aAb, \\ A &\rightarrow \lambda. \end{aligned}$$

Derive a few specific sentences to convince yourself that this works.

انواع گرامرها

$$P: \quad x \to y,$$

- Type 0
- Type 1
- Type 2
  CFG
  $y \in (V \cup T)^*$
  $x \in V$
- Type 3
  Regular

گرامر/زبان مستقل از متن (CFG/CFL)

DEFINITION 2.2

A ***context-free grammar*** is a 4-tuple (V, Σ, R, S), where

1. V is a finite set called the ***variables***,
2. Σ is a finite set, disjoint from V, called the ***terminals***,
3. R is a finite set of ***rules***, with each rule being a variable and a string of variables and terminals, and
4. $S \in V$ is the start variable.

A language L is said to be context-free if and only if there is a context-free grammar G such that $L = L(G)$.

گرامر مستقل از متن (CFG)

- دلیل نامگذاری

گرامر مستقل از متن (CFG)

An important application of context-free grammars occurs in the specification and compilation of programming languages. A grammar for a programming language often appears as a reference for people trying to learn the language syntax. Designers of compilers and interpreters for programming languages often start by obtaining a grammar for the language. Most compilers and interpreters contain a component called a ***parser*** that extracts the meaning of a program prior to generating the compiled code or performing the interpreted execution. A number of methodologies facilitate the construction of a parser once a context-free grammar is available. Some tools even automatically generate the parser from the grammar.

گرامر مستقل از متن (CFG)

- چه زبان‌هایی را توصیف میکند و برعکس؟؟؟
- رابطه با زبان‌های منظم؟
- آیا زبان مستقل از متن نیز تحت عملگرهای اجتماع، ستاره، الحاق و ... بسته است؟
- اتوماتای متناظر با این زبان چیست؟
- هم ارزی اتوماتای معین و نامعین؟
- محدودیت دارد؟

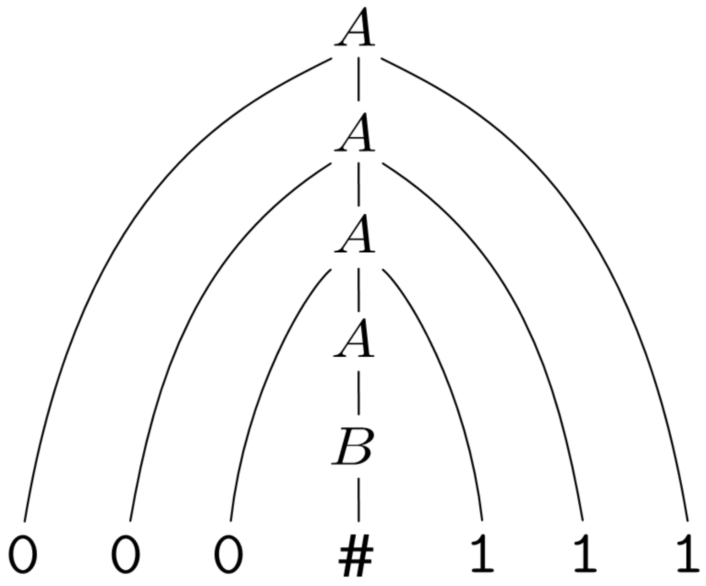
مثال

2.1

CONTEXT-FREE GRAMMARS

The following is an example of a context-free grammar, which we call G_1.

$$
\begin{aligned}
A &\rightarrow \mathtt{0}A\mathtt{1} \\
A &\rightarrow B \\
B &\rightarrow \mathtt{\#}
\end{aligned}
$$



Parse tree for 000#111 in grammar G_1

مثال

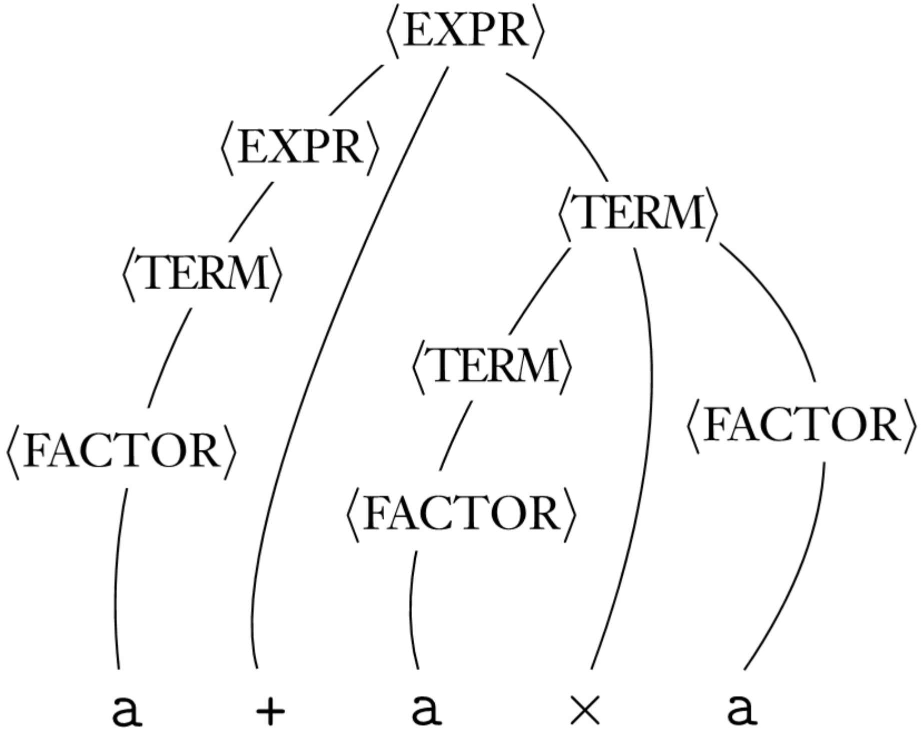
EXAMPLE 2.4

Consider grammar $G_4 = (V, \Sigma, R, \langle\text{EXPR}\rangle)$.

V is $\{\langle\text{EXPR}\rangle, \langle\text{TERM}\rangle, \langle\text{FACTOR}\rangle\}$ and Σ is $\{\mathtt{a}, \mathtt{+}, \mathtt{x}, \mathtt{(}, \mathtt{)}\}$. The rules are

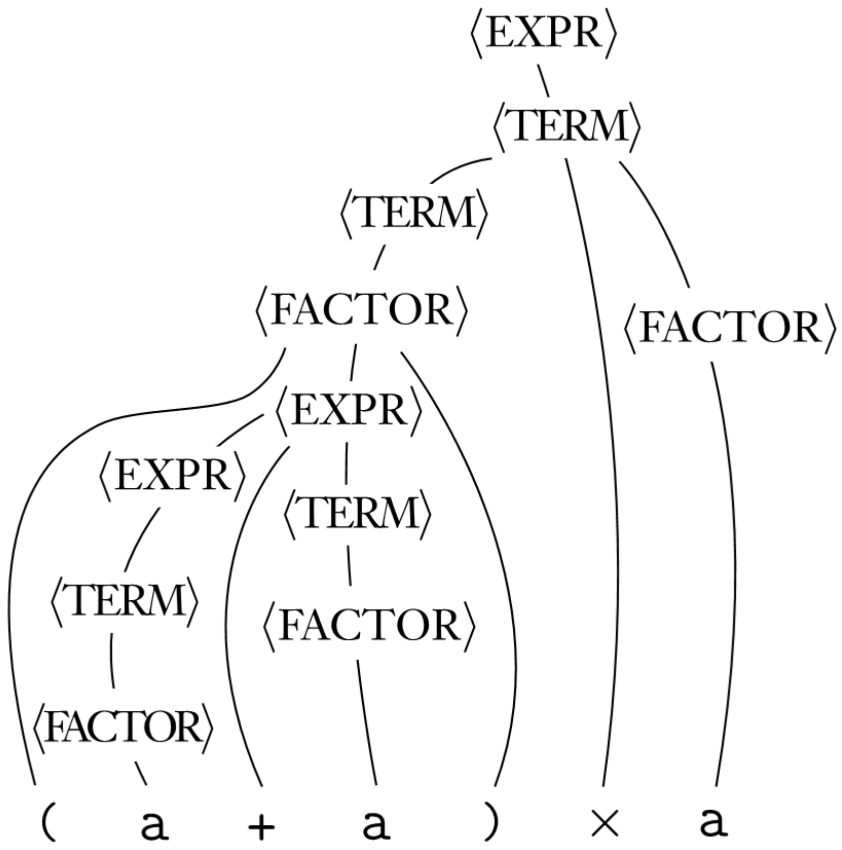
$$\begin{aligned}
\langle\text{EXPR}\rangle &\to \langle\text{EXPR}\rangle\mathtt{+}\langle\text{TERM}\rangle \mid \langle\text{TERM}\rangle \\
\langle\text{TERM}\rangle &\to \langle\text{TERM}\rangle\mathtt{x}\langle\text{FACTOR}\rangle \mid \langle\text{FACTOR}\rangle \\
\langle\text{FACTOR}\rangle &\to \mathtt{(}\,\langle\text{EXPR}\rangle\,\mathtt{)} \mid \mathtt{a}
\end{aligned}$$

مثال



L(G4)=language of arithmetic expressions

مثال



L(G4)=language of arithmetic expressions

مثال

EXAMPLE 5.1

The grammar $G = (\{S\}, \{a, b\}, S, P)$, with productions

$$S \rightarrow aSa,$$
$$S \rightarrow bSb,$$
$$S \rightarrow \lambda,$$

is context-free. A typical derivation in this grammar is

$$S \Rightarrow aSa \Rightarrow aaSaa \Rightarrow aabSbaa \Rightarrow aabbaa.$$

This, and similar derivations, make it clear that

$$L(G) = \left\{ww^R : w \in \{a, b\}^*\right\}.$$

The language is context-free, but as shown in Example 4.8, it is not regular.

مثال

$G = (\{A, B, S\}, \{a, b\}, S, P)$ with productions

1. $S \rightarrow AB.$
2. $A \rightarrow aaA.$
3. $A \rightarrow \lambda.$
4. $B \rightarrow Bb.$
5. $B \rightarrow \lambda.$

This grammar generates the language $L(G) = \{a^{2n}b^m : n \geq 0, m \geq 0\}.$

مثال

$G = (\{A, B, S\}, \{a, b\}, S, P)$ with productions

$$\mathbf{1}.\ S \rightarrow AB.$$
$$\mathbf{2}.\ A \rightarrow aaA.$$
$$\mathbf{3}.\ A \rightarrow \lambda.$$
$$\mathbf{4}.\ B \rightarrow Bb.$$
$$\mathbf{5}.\ B \rightarrow \lambda.$$

This grammar generates the language $L(G) = \{a^{2n}b^m : n \geq 0, m \geq 0\}$.

$$S \stackrel{1}{\Rightarrow} AB \stackrel{2}{\Rightarrow} aaAB \stackrel{3}{\Rightarrow} aaB \stackrel{4}{\Rightarrow} aaBb \stackrel{5}{\Rightarrow} aab$$

$$S \stackrel{1}{\Rightarrow} AB \stackrel{4}{\Rightarrow} ABb \stackrel{2}{\Rightarrow} aaABb \stackrel{5}{\Rightarrow} aaAb \stackrel{3}{\Rightarrow} aab.$$

تعریف

DEFINITION 5.2

A derivation is said to be **leftmost** if in each step the leftmost variable in the sentential form is replaced. If in each step the rightmost variable is replaced, we call the derivation **rightmost**.

مثال

IUT-ECE

EXAMPLE 5.5

Consider the grammar with productions

$$
\begin{aligned}
S &\to aAB, \\
A &\to bBb, \\
B &\to A|\lambda.
\end{aligned}
$$

Then

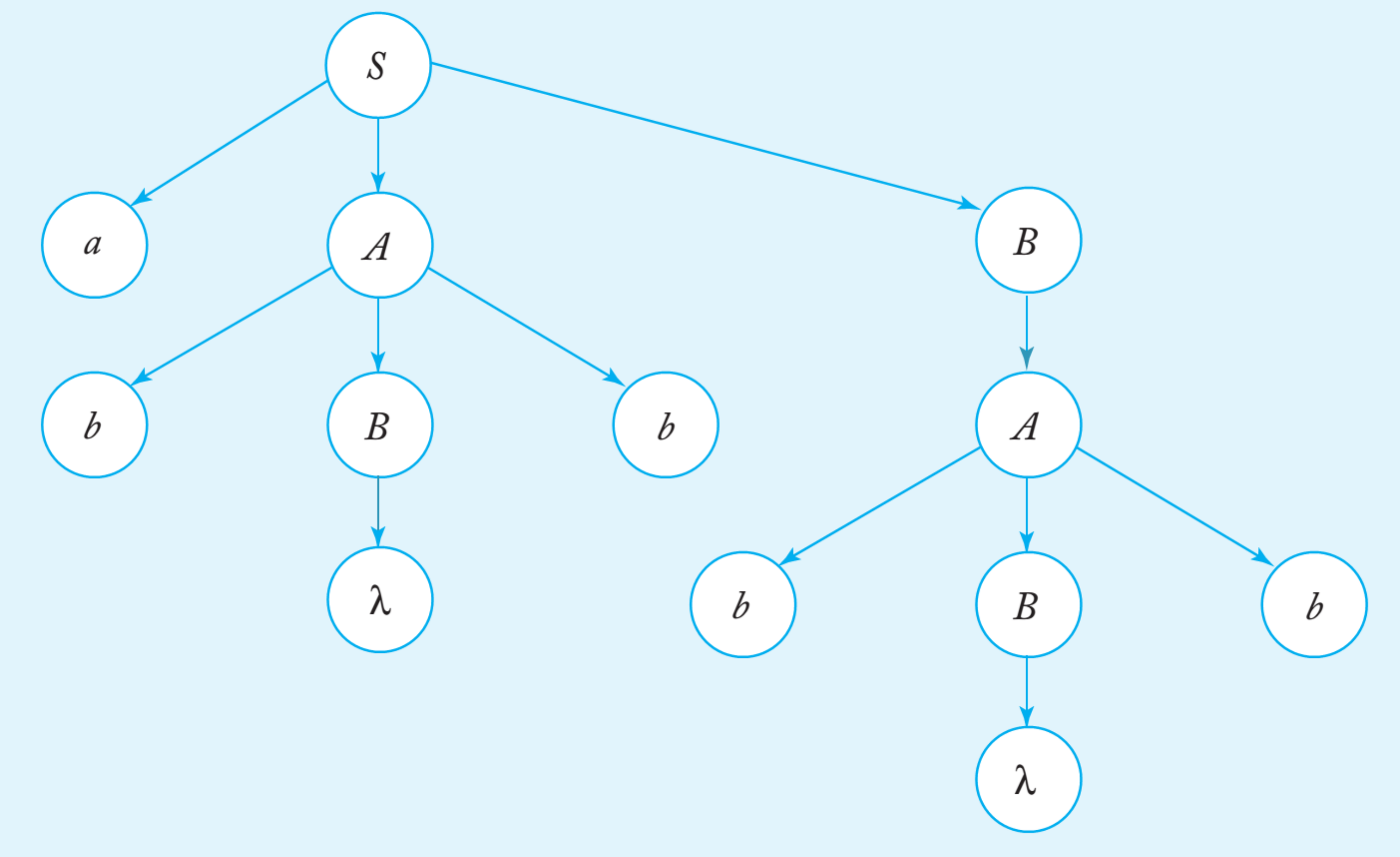
$$S \Rightarrow aAB \Rightarrow abBbB \Rightarrow abAbB \Rightarrow abbBbbB \Rightarrow abbbbB \Rightarrow abbbb$$

is a leftmost derivation of the string $abbbb$. A rightmost derivation of the same string is

$$S \Rightarrow aAB \Rightarrow aA \Rightarrow abBb \Rightarrow abAb \Rightarrow abbBbb \Rightarrow abbbb.$$

مثال

$$S \rightarrow aAB,$$
$$A \rightarrow bBb,$$
$$B \rightarrow A|\lambda.$$



مثال

EXAMPLE 2.3

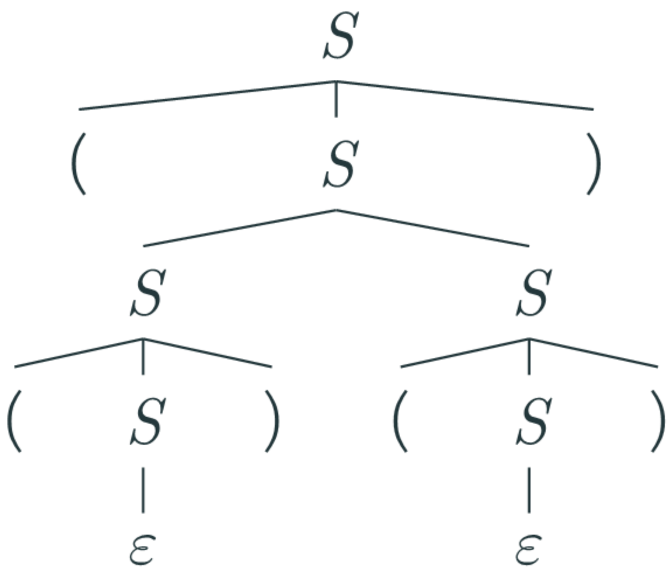
Consider grammar $G_3 = (\{S\}, \{\mathtt{a}, \mathtt{b}\}, R, S)$. The set of rules, R, is

$$S \rightarrow \mathtt{a}S\mathtt{b} \mid SS \mid \varepsilon.$$

This grammar generates strings such as abab, aaabbb, and aababb. You can see more easily what this language is if you think of a as a left parenthesis "(" and b as a right parenthesis ")". Viewed in this way, $L(G_3)$ is the language of all strings of properly nested parentheses. Observe that the right-hand side of a rule may be the empty string ε. ■

مثال

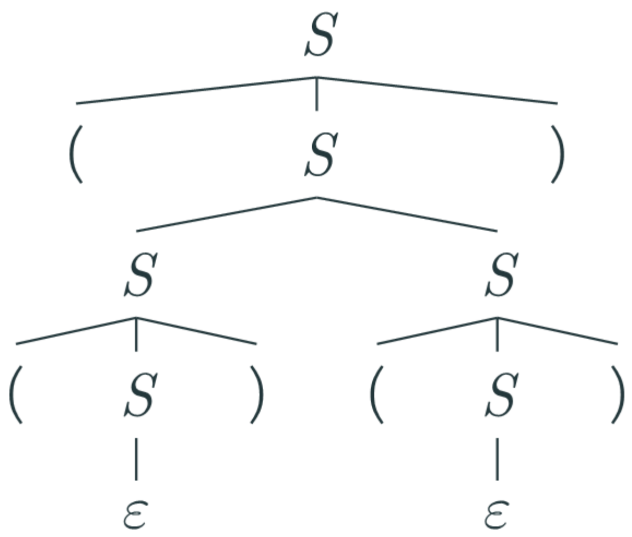
$$
\begin{aligned}
S &\Rightarrow (S) \\
&\Rightarrow (SS) \\
&\Rightarrow ((S)S) \\
&\Rightarrow ((S)(S)) \\
&\Rightarrow (()(S)) \\
&\Rightarrow (()())
\end{aligned}
$$



مثال

$$
\begin{aligned}
S &\Rightarrow (S) \\
&\Rightarrow (SS) \\
&\Rightarrow ((S)S) \\
&\Rightarrow ((S)(S)) \\
&\Rightarrow (()(S)) \\
&\Rightarrow (()())
\end{aligned}
$$

$$
\begin{aligned}
S &\Rightarrow (S) \\
&\Rightarrow (SS) \\
&\Rightarrow ((S)S) \\
&\Rightarrow (()S) \\
&\Rightarrow (()(S)) \\
&\Rightarrow (()())
\end{aligned}
$$



$$
\begin{aligned}
S &\Rightarrow (S) \\
&\Rightarrow (SS) \\
&\Rightarrow (S(S)) \\
&\Rightarrow ((S)(S)) \\
&\Rightarrow (()(S)) \\
&\Rightarrow (()())
\end{aligned}
$$

$$
\begin{aligned}
S &\Rightarrow (S) \\
&\Rightarrow (SS) \\
&\Rightarrow (S(S)) \\
&\Rightarrow (S()) \\
&\Rightarrow ((S)()) \\
&\Rightarrow (()())
\end{aligned}
$$