Overview

reg. languages	FA	reg. grammar	reg. expression
determ. cf. languages	DPDA		
cf. languages	PDA	cf. grammar	
cs. languages	LBA	cs. grammar	
re. languages	TM	unrestr. grammar	

From lecture 7:
$AnBn = \{ a^n b^n \mid n \geq 0 \} \subseteq \{a, b\}^*$

Example

– $\Lambda \in AnBn$ (basis)
– for every $x \in AnBn$, also $axb \in AnBn$ (induction)

$S \to \Lambda$
$S \to aSb$

$$S \Rightarrow aSb \Rightarrow aaSbb \Rightarrow aa\,bb$$
$$S \Rightarrow aSb \Rightarrow aaSbb \Rightarrow aaaSbbb \Rightarrow aaa\,bbb$$

if $S \Rightarrow^* x$ then also $S \Rightarrow^* axb$

From lecture 7:

Definition

context-free grammar (CFG) 4-tuple $G = (V, \Sigma, S, P)$
- V alphabet *variables* / *nonterminals*
- Σ alphabet *terminals* disjoint $V \cap \Sigma = \varnothing$
- $S \in V$ *axiom*, *start symbol*
- P finite set rules, *productions*
  of the form $A \to \alpha$, $A \in V$, $\alpha \in (V \cup \Sigma)^*$

derivation step $\alpha = \alpha_1 A \alpha_2 \Rightarrow_G \alpha_1 \gamma \alpha_2 = \beta$ for $A \to \gamma \in P$

Definition

language generated by G
$L(G) = \{ x \in \Sigma^* \mid S \Rightarrow_G^* x \}$

[M] Def 4.6 & 4.7

From lecture 7:

Using building blocks

Theorem

If L_1, L_2 are CFL, then so are $L_1 \cup L_2$, $L_1 L_2$ and L_1^*.

[M] Thm 4.9

Hence, CFL is closed onder union, concatenation, star

Regular languages are closed under
– Boolean operations (complement, union, intersection, minus)
– Regular operations (union, concatenation, star)
– Reverse (mirror)
– [inverse] Homomorphism

Fact, proof follows $\hookrightarrow$later

Theorem

the languages
– $AnBnCn = \{\ a^n b^n c^n \mid n \geq 0\ \}$ *and*
– $XX = \{xx \mid x \in \{a, b\}^*\ \}$
are not context-free

[M] E 6.3, E 6.4

AnBnCn is the intersection of two context-free languages
[M] E 6.10

The complement of both *AnBnCn* and *XX* is context-free.
[M] E 6.11
Hence, CFL is not closed under intersection, complement

$S \to S_1 \mid S_2$	union
$S \to S_1 S_2$	concatenation
$S \to S S_1 \mid \Lambda$	star

CFG for $\varnothing \dots$
CFG for $\{\sigma\} \dots$

Example

$L = bba(ab)^* + (ab + ba^* b)^* ba$

[M] E 4.11

Regular languages and CF grammars

$S \to S_1 \mid S_2$ union
$S \to S_1 S_2$ concatenation
$S \to SS_1 \mid \Lambda$ star

Example

$L = bba(ab)^* + (ab + ba^*b)^*ba$
$S \to S_1 \mid S_2$
$S_1 \to S_1 ab \mid bba$
$S_2 \to TS_2 \mid ba \quad T \to ab \mid bUb \quad U \to aU \mid \Lambda$

[M] E 4.11

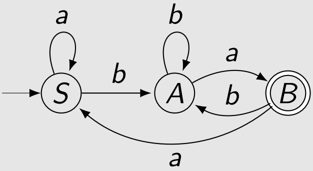
ABOVE

We have seen constructions to apply the regular operations (union, concatenation and star) to context-free grammars. These we can now use to build CFG for regular expressions.

There is a better way to build CFG for regular languages. Use finite automata, and simulate these using a very simple type of context-free grammar. These simple grammars are called regular.

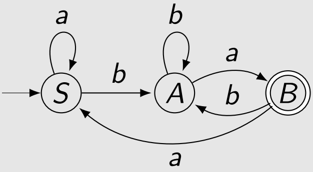
systematic approach

Example



systematic approach

Example



axiom S	initial state
$S \to bA \mid aS$	transitions
$A \to bA \mid aB$	
$B \to bA \mid aS$	
$B \to \Lambda$	accepting state

path / derivation for *bbaaba*...

Definition

regular grammar (or *right-linear grammar*)

productions are of the form

– $A \to \sigma B$ variables A, B, terminal σ

– $A \to \Lambda$ variable A

Special type of context-free grammar

Theorem

A language L is regular,
if and only if there is a regular grammar generating L.

Proof. . .

[M] Def 4.13, Thm 4.14

4.4 Derivation trees and ambiguity

A derivation. . .

$S \to a \mid S + S \mid S * S \mid (S) \quad \Sigma = \{a, +, *, (,)\}$

$S \Rightarrow S + \underline{S} \Rightarrow S + (\underline{S}) \Rightarrow S + (\underline{S} * S) \Rightarrow$
$\underline{S} + (a * S) \Rightarrow a + (a * \underline{S}) \Rightarrow a + (a * a)$

[M] E 4.2, Fig 4.15

Definition

A derivation in a context-free grammar is a *leftmost* derivation, if at each step, a production is applied to the leftmost variable-occurrence in the current string.
A *rightmost* derivation is defined similarly.

[M] D 4.16

derivation step $\alpha = \alpha_1 A \alpha_2 \Rightarrow_G \alpha_1 \gamma \alpha_2 = \beta$ for $A \to \gamma \in P$

The derivation step is *leftmost* iff $\alpha_1 \in \Sigma^*$

We write $\alpha \overset{\ell}{\Rightarrow} \beta$

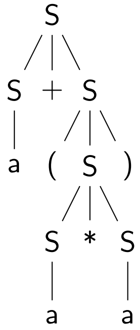
$S \to a \mid S + S \mid S * S \mid (S) \quad \Sigma = \{a, +, *, (,)\}$

$S \Rightarrow S + \underline{S} \Rightarrow S + (\underline{S}) \Rightarrow S + (\underline{S} * S) \Rightarrow$
$\underline{S} + (a * S) \Rightarrow a + (a * \underline{S}) \Rightarrow a + (a * a)$

Derivation tree. . .

[M] E 4.2, Fig 4.15

Expressions

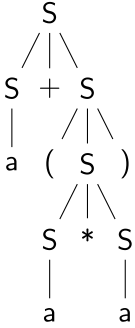


[M] E 4.2, Fig 4.15

$S \to a \mid S + S \mid S * S \mid (S) \quad \Sigma = \{a, +, *, (,)\}$

$S \Rightarrow S + \underline{S} \Rightarrow S + (\underline{S}) \Rightarrow S + (\underline{S} * S) \Rightarrow \underline{S} + (a * S) \Rightarrow a + (a * \underline{S}) \Rightarrow a + (a * a)$

Leftmost derivation. . .



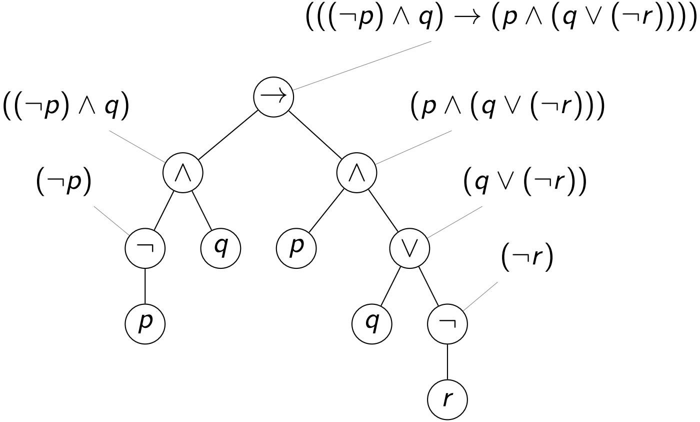
[M] E 4.2, Fig 4.15

$S \to a \mid S + S \mid S * S \mid (S) \quad \Sigma = \{a, +, *, (,)\}$

$S \Rightarrow S + \underline{S} \Rightarrow S + (\underline{S}) \Rightarrow S + (\underline{S} * S) \Rightarrow \underline{S} + (a * S) \Rightarrow a + (a * \underline{S}) \Rightarrow a + (a * a)$

Leftmost derivation:
$S \overset{\ell}{\Rightarrow} \underline{S} + S \overset{\ell}{\Rightarrow} a + S \overset{\ell}{\Rightarrow} a + (\underline{S}) \overset{\ell}{\Rightarrow} a + (\underline{S} * S) \overset{\ell}{\Rightarrow} a + (a * \underline{S}) \overset{\ell}{\Rightarrow} a + (a * a)$

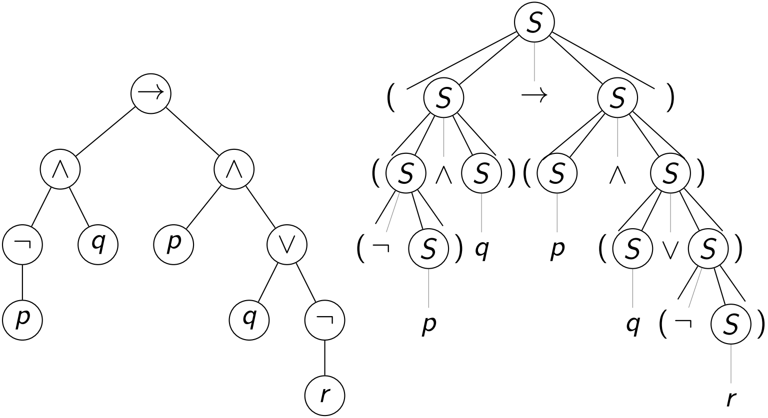
$\psi ::= p \mid (\neg\psi) \mid (\psi \wedge \psi) \mid (\psi \vee \psi) \mid (\psi \rightarrow \psi)$



[H&R] Fig 1.3

$S ::= p \mid q \mid r \mid (\neg S) \mid (S \wedge S) \mid (S \vee S) \mid (S \rightarrow S)$

parse tree vs. derivation tree[2]

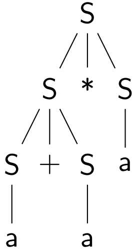


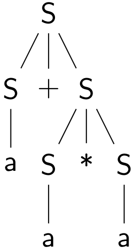
[2]with all brackets explicit

Definition

A context-free grammar G is *ambiguous*, if for at least one $x \in L(G)$, x has more than one derivation tree.

Otherwise: *unambiguous* [M] D 4.18





$\Sigma = \{a, +, *, (,)\}$
$S \rightarrow a \mid S + S \mid S * S \mid (S)$

$a + a * a$

leftmost derivation $\longleftrightarrow$ derivation tree

Theorem

If G is a context-free grammar, then for every $x \in L(G)$, these three statements are equivalent:

1. *x has more than one derivation tree*
2. *x has more than one leftmost derivation*
3. *x has more than one rightmost derivation*

Proof. . .

[M] Thm 4.17

leftmost derivation $\longleftrightarrow$ derivation tree

Theorem

If G is a context-free grammar, then for every $x \in L(G)$, these three statements are equivalent:

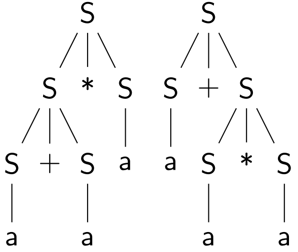
1. *x has more than one derivation tree*
2. *x has more than one leftmost derivation*
3. *x has more than one rightmost derivation*

[M] Thm 4.17

Definition

A context-free grammar G is *ambiguous*, if for at least one $x \in L(G)$, x has more than one derivation tree (or, equivalently, more than one leftmost derivation).

Otherwise: *unambiguous* [M] D 4.18



$$\Sigma = \{a, +, *, (,)\}$$
$$S \to a \mid S + S \mid S * S \mid (S)$$

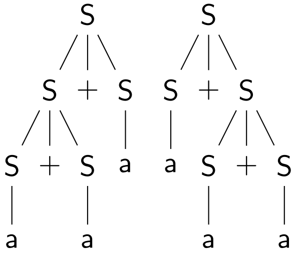
$a + a * a$

$$S \overset{\ell}{\Rightarrow} \underline{S} * S \overset{\ell}{\Rightarrow} S + S * S \overset{\ell}{\Rightarrow} a + S * S \overset{\ell}{\Rightarrow} a + a * S \overset{\ell}{\Rightarrow} a + a * a$$

$$S \overset{\ell}{\Rightarrow} \underline{S}{+}S \overset{\ell}{\Rightarrow} a{+}S \overset{\ell}{\Rightarrow} a{+}S{*}S \overset{\ell}{\Rightarrow} a{+}a{*}S \overset{\ell}{\Rightarrow} a + a * a$$

leftmost derivation $\longleftrightarrow$ derivation tree

Ambiguity (2)



$\Sigma = \{a, +, *, (,)\}$

$S \to a \mid S + S \mid S * S \mid (S)$

$a + a + a$

Leftmost for 1:

$S \overset{\ell}{\Rightarrow} \underline{S} + S \overset{\ell}{\Rightarrow} S + S + S \overset{\ell}{\Rightarrow} a + S + S \overset{\ell}{\Rightarrow} a + a + S \overset{\ell}{\Rightarrow} a + a + a$

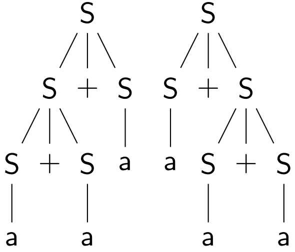
Derivation for 2:

$S \Rightarrow S + \underline{S} \Rightarrow S + S + S \Rightarrow a + S + S \Rightarrow a + a + S \Rightarrow a + a + a$

Ambiguity (2)

$\Sigma = \{a, +, *, (,)\}$
$S \to a \mid S + S \mid S * S \mid (S)$

$a + a + a$



Leftmost for 1:
$S \overset{\ell}{\Rightarrow} \underline{S} + S \overset{\ell}{\Rightarrow} S + S + S \overset{\ell}{\Rightarrow} a + S + S \overset{\ell}{\Rightarrow} a + a + S \overset{\ell}{\Rightarrow} a + a + a$

Derivation for 2:
$S \Rightarrow S + \underline{S} \Rightarrow S + S + S \Rightarrow a + S + S \Rightarrow a + a + S \Rightarrow a + a + a$

Leftmost for 2:
$S \overset{\ell}{\Rightarrow} \underline{S} + S \overset{\ell}{\Rightarrow} a + S \overset{\ell}{\Rightarrow} a + S + S \overset{\ell}{\Rightarrow} a + a + S \overset{\ell}{\Rightarrow} a + a + a$

leftmost derivation $\longleftrightarrow$ derivation tree

This example is a little weird. In the derivation step $\mathsf{S} + \mathsf{S} \Rightarrow \mathsf{S} + \mathsf{S} + \mathsf{S}$ we cannot really see which S has been rewritten.

Expr

ambiguous:

$S \to a \mid S + S \mid S * S \mid (S)$

[M] E 4.20

$a + a * a$

unambiguous:

...

Expr

ambiguous:

$S \to a \mid S + S \mid S * S \mid (S)$

[M] E 4.20

$a + a * a$

unambiguous:

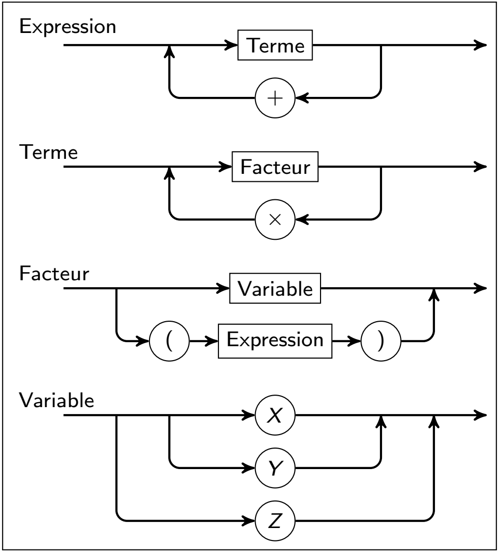
$S \to S + T \mid T$

$T \to T * F \mid F$

$F \to a \mid (S)$

[M] Thm 4.25

The proof of the unambiguity does not have to be known for the exam



right associative

http://math.et.info.free.fr/TikZ/index.html Chapitre 7

$AeqB = \{ x \in \{a, b\}^* \mid n_a(x) = n_b(x) \}$

$aaabbb, ababab, aababb, \ldots$

$S \to \Lambda \mid aB \mid bA$
$A \to aS \mid bAA$ — A generates $n_a(x) = n_b(x) + 1$
$B \to bS \mid aBB$ — B generates $n_a(x) + 1 = n_b(x)$

Derivation for *aababb*:

$S \Rightarrow aB \Rightarrow aaBB \Rightarrow aabSB \Rightarrow \ldots$ (different options)

(1) $aabB \Rightarrow aabaBB \Rightarrow aababSB \Rightarrow aababB \Rightarrow aababbS \Rightarrow aababb$
(2) $aaba\underline{B}B \Rightarrow aabab\underline{S}B \Rightarrow aabab\underline{B} \Rightarrow aababb\underline{S} \Rightarrow aababb$
(2') $aabaB\underline{B} \Rightarrow aaba\underline{B}bS \Rightarrow aababSb\underline{S} \Rightarrow aabab\underline{S}b \Rightarrow aababb$

[M] E 4.8

ABOVE

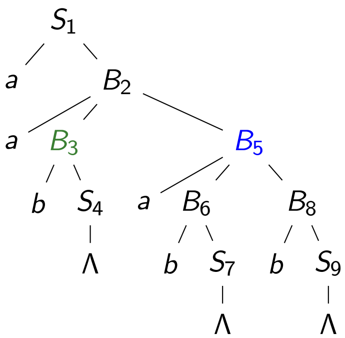
When a string has multiple variables, like $aabSB$ in the above example, then we are not forced to rewrite the first variable, we can as well rewrite another one.

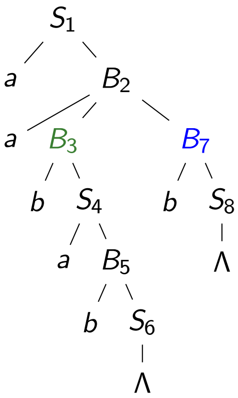
Thus we can do $aab\underline{S}B \Rightarrow aabB$, but also $aabS\underline{B} \Rightarrow aabSaBB$, for instance.

BELOW

In detail, two different derivation trees for the same string, corresponding to derivations (1) and (2,2') respectively, together with two associated leftmost derivations.

Given these two trees we conclude the grammar is ambiguous.





$S \Rightarrow aB \Rightarrow aaBB \Rightarrow aabSB \Rightarrow$
$aabB \Rightarrow aabaBB \Rightarrow aababSB \Rightarrow$
$aababB \Rightarrow aababbS \Rightarrow aababb$

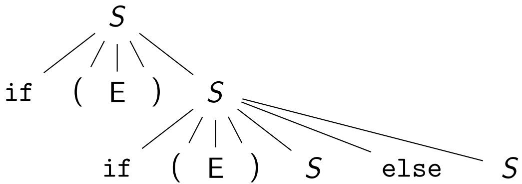
$S \Rightarrow aB \Rightarrow aaBB \Rightarrow aabSB \Rightarrow$
$aabaBB \Rightarrow aababSB \Rightarrow aababB \Rightarrow$
$aababbS \Rightarrow aababb$

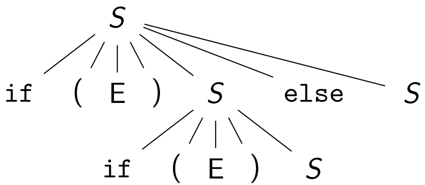
$S \to \texttt{if} \; (\; E \;) \; S \;\mid\; \texttt{if} \; (\; E \;) \; S \; \texttt{else} \; S \;\mid\; \ldots$

$\texttt{if} \; (\; E \;) \; \texttt{if} \; (\; E \;) \; S \; \texttt{else} \; S$

[M] E 4.19

$S \to$ if (E) S | if (E) S else S | ...





[M] E 4.19

ambiguous:

$S \to \texttt{if}\ (\ E\ )\ S \mid \texttt{if}\ (\ E\ )\ S\ \texttt{else}\ S \mid A \mid \ \ldots$

unambiguous...

[M] E 4.19

ambiguous:

$S \to \texttt{if}\ (\ E\ )\ S \mid \texttt{if}\ (\ E\ )\ S\ \texttt{else}\ S \mid A \mid \ \dots$

unambiguous:

$S \to S_1 \mid S_2$

$S_1 \to \texttt{if}\ (\ E\ )\ S_1\ \texttt{else}\ S_1 \mid A \mid \ \dots$ (matched)

$S_2 \to \texttt{if}\ (\ E\ )\ S \mid \texttt{if}\ (\ E\ )\ S_1\ \texttt{else}\ S_2$ (open)

[M] E 4.19

Balanced

ambiguous:

$S \to SS \mid (S) \mid \Lambda$ (more or less the definition of balanced)

unambiguous:

$S \to (S)S \mid \Lambda$

[M] Exercise 4.45

Some cf languages are inherently ambiguous

Ambiguity is *undecidable*

[M] Theorem 9.20

Let G be a context-free grammar with start variable S and the following productions:

$$S \rightarrow aSbS \mid bSaS \mid \Lambda$$

a. Show that $L(G) = AEqB = \{x \in \{a, b\}^* \mid n_a(x) = n_b(x)\}$
b. Is G ambiguous? Motivate your answer.