

4.4 Derivation trees and ambiguity

A derivation...

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

$$\begin{aligned} S &\Rightarrow S + \underline{S} \Rightarrow S + (\underline{S}) \Rightarrow S + (\underline{S} * S) \Rightarrow \\ \underline{S} + (a * S) &\Rightarrow a + (a * \underline{S}) \Rightarrow a + (a * a) \end{aligned}$$

[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 $A \rightarrow \gamma \in P$

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

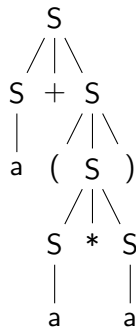
We write $\alpha \xRightarrow{\ell} \beta$

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

$$\begin{aligned} S &\Rightarrow S + \underline{S} \Rightarrow S + (\underline{S}) \Rightarrow S + (\underline{S} * S) \Rightarrow \\ \underline{S} + (a * S) &\Rightarrow a + (a * \underline{S}) \Rightarrow a + (a * a) \end{aligned}$$

Derivation tree...

[M] E 4.2, Fig 4.15

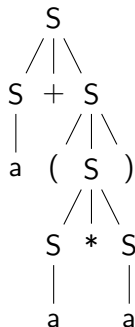


$$S \rightarrow a \mid S + S \mid S * S \mid (S) \quad \Sigma = \{a, +, *, (,)\}$$

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

Leftmost derivation...

[M] E 4.2, Fig 4.15



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

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

Leftmost derivation:

$$S \xRightarrow{\ell} \underline{S} + S \xRightarrow{\ell} a + S \xRightarrow{\ell} a + (\underline{S}) \xRightarrow{\ell} a + (\underline{S} * S) \xRightarrow{\ell} a + (a * \underline{S}) \xRightarrow{\ell} a + (a * a)$$

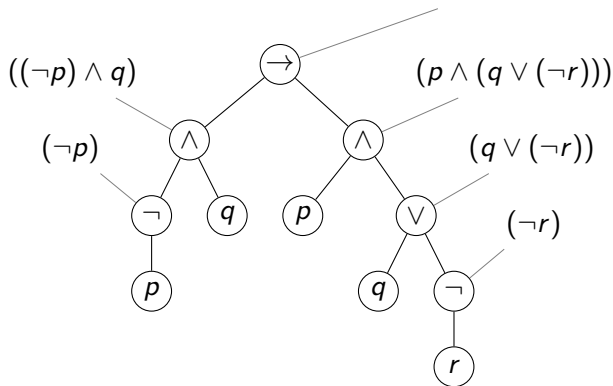
One tree $\leftrightarrow$ multiple derivations

Tree $\rightarrow$ derivation $\rightarrow$ (same) tree

[M] E 4.2, Fig 4.15

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

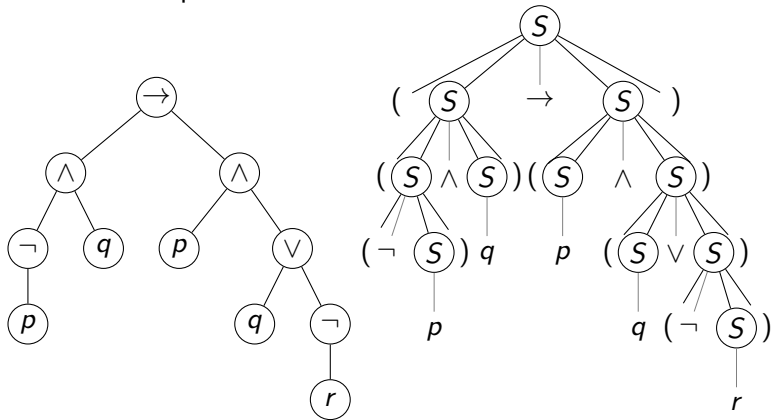
$$(((\neg p) \wedge q) \rightarrow (p \wedge (q \vee (\neg r))))$$



[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 tree¹

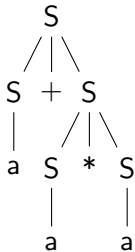
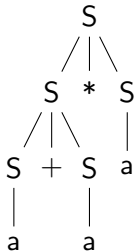


¹with all brackets explicit

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

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

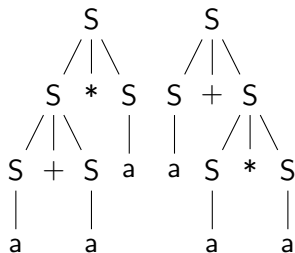
$$S \rightarrow a \mid S + S \mid S * S \mid (S)$$

$$a + a * a$$

$$S \xRightarrow{\ell} \underline{S} * S \xRightarrow{\ell} S + S * S \xRightarrow{\ell} a + S * S \xRightarrow{\ell} a + a * S \xRightarrow{\ell} a + a * a$$

$$S \xRightarrow{\ell} \underline{S} + S \xRightarrow{\ell} a + S \xRightarrow{\ell} a + S * S \xRightarrow{\ell} a + a * S \xRightarrow{\ell} a + a * a$$

leftmost derivation $\longleftrightarrow$ derivation tree



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

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

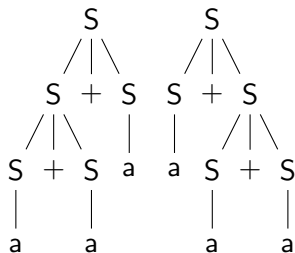
$$a + a + a$$

Leftmost for 1:

$$\begin{aligned} S &\Rightarrow \underline{S} + S \xRightarrow{\ell} S + S + S \xRightarrow{\ell} a + S + S \xRightarrow{\ell} \\ &a + a + S \xRightarrow{\ell} a + a + a \end{aligned}$$

Derivation for 2:

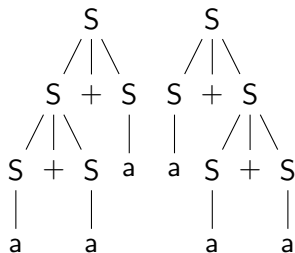
$$\begin{aligned} S &\Rightarrow S + \underline{S} \Rightarrow S + S + S \Rightarrow a + S + S \Rightarrow \\ &a + a + S \Rightarrow a + a + a \end{aligned}$$



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

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

$$a + a + a$$



Leftmost for 1:

$$S \xRightarrow{\ell} \underline{S} + S \xRightarrow{\ell} S + S + S \xRightarrow{\ell} a + S + S \xRightarrow{\ell} a + a + S \xRightarrow{\ell} a + a + a$$

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:

$$S \xRightarrow{\ell} \underline{S} + S \xRightarrow{\ell} a + S \xRightarrow{\ell} a + S + S \xRightarrow{\ell} a + a + S \xRightarrow{\ell} a + a + a$$

leftmost derivation $\longleftrightarrow$ derivation tree

ABOVE

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

Expr

ambiguous:

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

[M] E 4.20

$a + a * a$

unambiguous:

...

Expr

ambiguous:

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

[M] E 4.20

$a + a * a$

unambiguous:

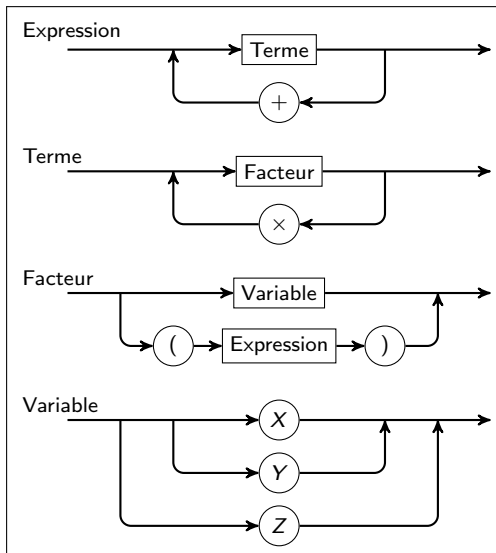
$S \rightarrow S + T \mid T$

$T \rightarrow T * F \mid F$

$F \rightarrow a \mid (S)$

[M] Thm 4.25

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



right associative

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

aaabbb, ababab, aababb, ...

$$S \rightarrow \Lambda \mid aB \mid bA$$

$$A \rightarrow aS \mid bAA$$

$$B \rightarrow bS \mid aBB$$

A generates $n_a(x) = n_b(x) + 1$

B generates $n_a(x) + 1 = n_b(x)$

Derivation for *aababb*:

$$S \Rightarrow aB \Rightarrow aa\textcolor{teal}{B}B \Rightarrow aa\textcolor{teal}{b}S\textcolor{blue}{B} \Rightarrow \dots \quad (\text{different options})$$

$$(1) aa\textcolor{teal}{b}B \Rightarrow aa\textcolor{teal}{b}aBB \Rightarrow aa\textcolor{teal}{b}abSB \Rightarrow aa\textcolor{teal}{b}abB \Rightarrow aa\textcolor{teal}{b}abbS \Rightarrow aa\textcolor{teal}{b}abb$$

$$(2) aa\textcolor{teal}{b}a\underline{B}B \Rightarrow aa\textcolor{teal}{b}ab\underline{S}B \Rightarrow aa\textcolor{teal}{b}ab\underline{B} \Rightarrow aa\textcolor{teal}{b}abb\underline{S} \Rightarrow aa\textcolor{teal}{b}abb$$

$$(2') aa\textcolor{teal}{b}a\underline{B}B \Rightarrow aa\textcolor{teal}{b}a\underline{B}bS \Rightarrow aa\textcolor{teal}{b}abS\underline{b}S \Rightarrow aa\textcolor{teal}{b}ab\underline{S}b \Rightarrow aa\textcolor{teal}{b}abb$$

[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.

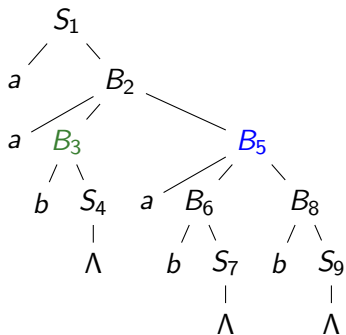
Thus we can do $aab\underline{S}B \Rightarrow aabB$, but also $aabS\underline{B} \Rightarrow aabSaBB$, for instance.

BELOW

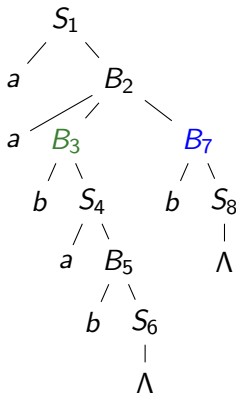
In detail, two different derivation trees for the same string, corresponding to derivations (1) and (2,2') respectively, together with two associated leftmost derivations.

Given these two trees we conclude the grammar is ambiguous.

Derivation tree & leftmost derivations



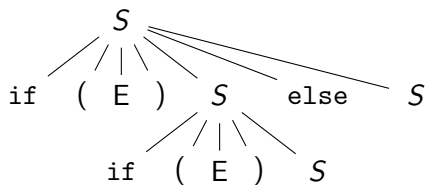
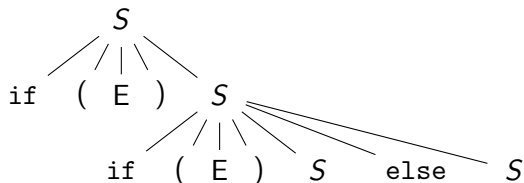
$S \Rightarrow aB \Rightarrow aaBB \Rightarrow aabSB \Rightarrow$
 $aabB \Rightarrow aabaBB \Rightarrow aababSB \Rightarrow$
 $aababbB \Rightarrow aababbS \Rightarrow aababb$



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

$$S \rightarrow \text{if } (E) S \mid \text{if } (E) S \text{ else } S \mid \dots$$
$$\text{if } (E) \text{if } (E) S \text{ else } S$$

[M] E 4.19

$$S \rightarrow \text{if } (E) S \mid \text{if } (E) S \text{ else } S \mid \dots$$


[M] E 4.19

ambiguous:

$S \rightarrow \text{if } (E) S \mid \text{if } (E) S \text{ else } S \mid A \mid \dots$

unambiguous...

[M] E 4.19

ambiguous:

$$S \rightarrow \text{if } (E) S \mid \text{if } (E) S \text{ else } S \mid A \mid \dots$$
unambiguous:

$$S \rightarrow S_1 \mid S_2$$

$$S_1 \rightarrow \text{if } (E) S_1 \text{ else } S_1 \mid A \mid \dots \quad (\text{matched})$$

$$S_2 \rightarrow \text{if } (E) S \mid \text{if } (E) S_1 \text{ else } S_2 \quad (\text{open})$$

[M] E 4.19

Balanced

ambiguous:

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

unambiguous:

$S \rightarrow (S)S \mid \Lambda$

[M] Exercise 4.45

Some cf languages are *inherently ambiguous*

Ambiguity is *undecidable*

[M] Theorem 9.20

unwanted in CFG:

– variables not used in successful derivations $S \Rightarrow^* x \in \Sigma^*$

CFG $G = (V, \Sigma, S, P)$

Definition

variable A is *live* if $A \Rightarrow^* x$ for some $x \in \Sigma^*$.

variable A is *reachable* if $S \Rightarrow^* \alpha A \beta$ for some $\alpha, \beta \in (\Sigma \cup V)^*$.

variable A is *useful* if there is a derivation of the form $S \Rightarrow^* \alpha A \beta \Rightarrow^* x$ for some string $x \in \Sigma^*$.

useful implies live and reachable.

conversely, ...

[M] Exercise 4.51, 4.52, 4.53

CFG $G = (V, \Sigma, S, P)$

Definition

variable A is *live* if $A \Rightarrow^* x$ for some $x \in \Sigma^*$.

variable A is *reachable* if $S \Rightarrow^* \alpha A \beta$ for some $\alpha, \beta \in (\Sigma \cup V)^*$.

variable A is *useful* if there is a derivation of the form $S \Rightarrow^* \alpha A \beta \Rightarrow^* x$ for some string $x \in \Sigma^*$.

useful implies live and reachable.

For $S \rightarrow AB \mid b$ and $A \rightarrow a$, variable A is live and reachable, not useful.

[M] Exercise 4.51, 4.52, 4.53

Live variables

Construction

- $N_0 = \emptyset$
- $N_{i+1} = N_i \cup \{ A \in V \mid A \rightarrow \alpha \text{ in } P, \text{ with } \alpha \in (N_i \cup \Sigma)^* \}$

$$N_1 = \{ A \in V \mid A \rightarrow x \text{ in } P, \text{ with } x \in \Sigma^* \}$$

$$N_0 \subseteq N_1 \subseteq N_2 \subseteq \dots \subseteq V$$

there exists a k such that $N_k = N_{k+1}$

A is **live** iff $A \in \bigcup_{i \geq 0} N_i = N_k$

(minimal) depth of derivation tree $A \Rightarrow^* x$

Live variables

Construction

- $N_0 = \emptyset$
- $N_{i+1} = N_i \cup \{ A \in V \mid A \rightarrow \alpha \text{ in } P, \text{ with } \alpha \in (N_i \cup \Sigma)^* \}$

Exercise 4.53(c.i).

$$S \rightarrow ABC \mid BaB$$

$$A \rightarrow aA \mid BaC \mid aaa$$

$$B \rightarrow bBb \mid a$$

$$C \rightarrow CA \mid AC$$

Reachable variables

Construction

- $N_0 = \{S\}$
- $N_{i+1} = N_i \cup \{ A \in V \mid B \rightarrow \alpha_1 A \alpha_2 \text{ in } P, \text{ with } B \in N_i \}$

$$N_0 \subseteq N_1 \subseteq N_2 \subseteq \dots \subseteq V$$

there exists a k such that $N_k = N_{k+1}$

A is **reachable** iff $A \in \bigcup_{i \geq 0} N_i = N_k$

(minimal) length of derivation $S \Rightarrow^* \alpha A \beta$

Reachable variables

Construction

- $N_0 = \{S\}$
- $N_{i+1} = N_i \cup \{ A \in V \mid B \rightarrow \alpha_1 A \alpha_2 \text{ in } P, \text{ with } B \in N_i \}$

$$N_0 \subseteq N_1 \subseteq N_2 \subseteq \dots \subseteq V$$

there exists a k such that $N_k = N_{k+1}$

A is **reachable** iff $A \in \bigcup_{i \geq 0} N_i = N_k$

(minimal) length of derivation $S \Rightarrow^* \alpha A \beta$

- remove all non-live variables (and productions that contain them)
- remove all unreachable variables (and their productions)

then all variables are useful

Reachable variables

Construction

- $N_0 = \{S\}$
- $N_{i+1} = N_i \cup \{ A \in V \mid B \rightarrow \alpha_1 A \alpha_2 \text{ in } P, \text{ with } B \in N_i \}$

Exercise 4.53(c.i)., ctd

$$S \rightarrow BaB$$

$$A \rightarrow aA \mid aaa$$

$$B \rightarrow bBb \mid a$$

- remove all non-live variables (and productions that contain them)
- remove all unreachable variables (and productions)

then all variables are useful

does not work the other way around ...

Exercise 4.53(c.i)., revisited

$$\begin{array}{ll}
 S \rightarrow ABC \mid BaB & A \rightarrow aA \mid BaC \mid aaa \\
 B \rightarrow bBb \mid a & C \rightarrow CA \mid AC
 \end{array}$$