

From lecture 8:

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

Some cf languages are *inherently ambiguous*

Ambiguity is *undecidable*

[M] Theorem 9.20

unwanted in CFG:

- variables not used in successful derivations $S \Rightarrow^* x \in \Sigma^*$
- $A \rightarrow \Lambda$ A variable Λ -productions
- $A \rightarrow B$ A, B variables unit productions [chain rules]

unwanted in CFG:

- variables not used in successful derivations $S \Rightarrow^* x \in \Sigma^*$
- $A \rightarrow \Lambda$ A variable Λ -productions
- $A \rightarrow B$ A, B variables unit productions [chain rules]

restricted CFG, with 'nice' form

Chomsky normalform $A \rightarrow BC, A \rightarrow \sigma$

Greibach normalform (\boxtimes) $A \rightarrow \sigma B_1 \dots B_k$

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).

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

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

does not work the other way around ...

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

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

Idea:

Example

$A \rightarrow BCDCB$

$B \rightarrow b \mid \Lambda$

$C \rightarrow c \mid \Lambda$

$D \rightarrow d$

Definition

variable A is *nullable* iff $A \Rightarrow^* \Lambda$

Theorem

- if $A \rightarrow \Lambda$ then A is nullable
- if $A \rightarrow B_1 B_2 \dots B_k$ and all B_i are nullable, then A is nullable

[M] Def 4.26 / Exercise 4.48

Construction

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

$$N_1 = \{ A \in V \mid A \rightarrow \Lambda \text{ in } P \}$$

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

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

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

Construction

- identify nullable variables
- for every production $A \rightarrow \alpha$ add $A \rightarrow \beta$,
where β is obtained from α by removing one or more nullable variables
- remove all Λ -productions (and all productions $A \rightarrow A$)

Grammar for $\{ a^i b^j c^k \mid i = j \text{ or } i = k \}$

$$S \rightarrow TU \mid V$$

$$T \rightarrow aTb \mid \Lambda$$

$$U \rightarrow cU \mid \Lambda$$

$$V \rightarrow aVc \mid W$$

$$W \rightarrow bW \mid \Lambda$$

Grammar for $\{ a^i b^j c^k \mid i = j \text{ or } i = k \}$

$$S \rightarrow TU \mid V$$

$$T \rightarrow aTb \mid \Lambda$$

$$U \rightarrow cU \mid \Lambda$$

$$V \rightarrow aVc \mid W$$

$$W \rightarrow bW \mid \Lambda$$

$N_1 = \{T, U, W\}$, variables with Λ at right-hand side productions

$N_2 = \{T, U, W\} \cup \{S, V\}$, variables with $\{T, U, W\}^*$ at rhs productions

$N_3 = N_2 = \{T, U, W, S, V\}$, all productions found, no new

add all productions, where (any number of) nullable variables are removed. . .

$$S \rightarrow TU \mid V$$

$$T \rightarrow aTb \mid \Lambda$$

$$U \rightarrow cU \mid \Lambda$$

$$V \rightarrow aVc \mid W$$

$$W \rightarrow bW \mid \Lambda$$

[M] Ex. 4.31

add all productions, where (any number of) nullable variables are removed

$$\begin{array}{ll}
 S \rightarrow TU \mid V & S \rightarrow T \mid U \mid \Lambda \\
 T \rightarrow aTb \mid \Lambda & T \rightarrow ab \\
 U \rightarrow cU \mid \Lambda & U \rightarrow c \\
 V \rightarrow aVc \mid W & V \rightarrow ac \mid \Lambda \\
 W \rightarrow bW \mid \Lambda & W \rightarrow b
 \end{array}$$

remove all Λ -productions. . .

[M] Ex. 4.31

add all productions, where (any number of) nullable variables are removed

$$\begin{array}{ll}
 S \rightarrow TU \mid V & S \rightarrow T \mid U \mid \Lambda \\
 T \rightarrow aTb \mid \Lambda & T \rightarrow ab \\
 U \rightarrow cU \mid \Lambda & U \rightarrow c \\
 V \rightarrow aVc \mid W & V \rightarrow ac \mid \Lambda \\
 W \rightarrow bW \mid \Lambda & W \rightarrow b
 \end{array}$$

remove all Λ -productions

$$\begin{array}{l}
 S \rightarrow TU \mid V \mid T \mid U \\
 T \rightarrow aTb \mid ab \\
 U \rightarrow cU \mid c \\
 V \rightarrow aVc \mid W \mid ac \\
 W \rightarrow bW \mid b
 \end{array}$$

[M] Ex. 4.31

Theorem

For every CFG G there is CFG G_1 without Λ -productions such that $L(G_1) = L(G) - \{\Lambda\}$.

Proof...

[M] Thm 4.27

Idea:

Example

$$A \rightarrow B \mid aCb$$
$$B \rightarrow C \mid Bb \mid Bc$$
$$C \rightarrow c \mid ABC$$

Assume Λ -productions have been removed

Variable B is *A-derivable*, if

- $B \neq A$, and
- $A \Rightarrow^* B$ (using only unit productions)

Construction

- $N_1 = \{ B \in V \mid B \neq A \text{ and } A \rightarrow B \text{ in } P \}$
- $N_{i+1} = N_i \cup \{ C \in V \mid C \neq A \text{ and } B \rightarrow C \text{ in } P, \text{ with } B \in N_i \}$

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

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

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

Construction

- for each $A \in V$, identify A -derivable variables
- for every pair (A, B) where B is A -derivable, and every production $B \rightarrow \alpha$ add $A \rightarrow \alpha$
- remove all unit productions

Grammar for $\{ a^i b^j c^k \mid i = j \text{ or } i = k \}$

$$S \rightarrow TU \mid V \mid T \mid U$$

$$T \rightarrow aTb \mid ab$$

$$U \rightarrow cU \mid c$$

$$V \rightarrow aVc \mid W \mid ac$$

$$W \rightarrow bW \mid b$$

$$S \rightarrow TU \mid V \mid T \mid U$$

$$T \rightarrow aTb \mid ab$$

$$U \rightarrow cU \mid c$$

$$V \rightarrow aVc \mid W \mid ac$$

$$W \rightarrow bW \mid b$$

S-derivable: $\{V, T, U\}, \{V, T, U, W\}$

V-derivable: $\{W\}$

New productions:

$$S \rightarrow aTb \mid ab \quad S \rightarrow cU \mid c \quad S \rightarrow aVc \mid W \mid ac \quad S \rightarrow bW \mid b$$

$$V \rightarrow bW \mid b$$

Remove unit productions:

$$S \rightarrow TU \mid aTb \mid ab \mid cU \mid c \mid aVc \mid ac \mid bW \mid b$$

$$T \rightarrow aTb \mid ab$$

$$U \rightarrow cU \mid c$$

$$V \rightarrow aVc \mid ac \mid bW \mid b$$

$$W \rightarrow bW \mid b$$

Definition

CFG in *Chomsky normal form*

productions are of the form

– $A \rightarrow BC$ variables A, B, C

– $A \rightarrow \sigma$ variable A , terminal σ

Theorem

For every CFG G there is CFG G_1 in CNF such that $L(G_1) = L(G) - \{\Lambda\}$.

[M] Def 4.29, Thm 4.30

Construction

- ① remove Λ -productions
- ② remove unit productions
- ③ introduce variables for terminals $X_\sigma \rightarrow \sigma$
- ④ split long productions

$$A \rightarrow aBabA$$

is replaced by

$$X_a \rightarrow a \quad X_b \rightarrow b \quad A \rightarrow X_a B X_a X_b A$$

$$A \rightarrow ACBA$$

is replaced by

$$A \rightarrow AY_1 \quad Y_1 \rightarrow CY_2 \quad Y_2 \rightarrow BA$$

Grammar for $\{ a^i b^j c^k \mid i = j \text{ or } i = k \}$

$$S \rightarrow TU \mid V$$

$$T \rightarrow aTb \mid \Lambda \quad U \rightarrow cU \mid \Lambda$$

$$V \rightarrow aVc \mid W \quad W \rightarrow bW \mid \Lambda$$

After removing Λ -productions and unit productions, we obtain (see before)

$$S \rightarrow TU \mid aTb \mid ab \mid cU \mid c \mid aVc \mid ac \mid bW \mid b$$

$$T \rightarrow aTb \mid ab \quad U \rightarrow cU \mid c$$

$$V \rightarrow aVc \mid ac \mid bW \mid b \quad W \rightarrow bW \mid b$$

Now introduce productions for the terminals...

Grammar for $\{ a^i b^j c^k \mid i = j \text{ or } i = k \}$

$$S \rightarrow TU \mid V$$

$$T \rightarrow aTb \mid \Lambda \quad U \rightarrow cU \mid \Lambda$$

$$V \rightarrow aVc \mid W \quad W \rightarrow bW \mid \Lambda$$

After removing Λ -productions and unit productions, we obtain (see before)

$$S \rightarrow TU \mid aTb \mid ab \mid cU \mid c \mid aVc \mid ac \mid bW \mid b$$

$$T \rightarrow aTb \mid ab \quad U \rightarrow cU \mid c$$

$$V \rightarrow aVc \mid ac \mid bW \mid b \quad W \rightarrow bW \mid b$$

Now introduce productions for the terminals:

$$X_a \rightarrow a \quad X_b \rightarrow b \quad X_c \rightarrow c$$

$$S \rightarrow TU \mid X_a T X_b \mid X_a X_b \mid X_c U \mid c \mid X_a V X_c \mid X_a X_c \mid X_b W \mid b$$

$$T \rightarrow X_a T X_b \mid X_a X_b$$

$$U \rightarrow X_c U \mid c$$

$$V \rightarrow X_a V X_c \mid X_a X_c \mid X_b W \mid b$$

$$W \rightarrow X_b W \mid b$$

Only a few productions that are too long:

$$S \rightarrow X_a T X_b \mid X_a V X_c$$

$$T \rightarrow X_a T X_b$$

$$V \rightarrow X_a V X_c$$

Split these long productions...

Only a few productions that are too long:

$$S \rightarrow X_a TX_b \mid X_a VX_c$$

$$T \rightarrow X_a TX_b$$

$$V \rightarrow X_a VX_c$$

Split these long productions:

$$S \rightarrow X_a Y_1 \mid X_a Y_2$$

$$Y_1 \rightarrow TX_b \quad Y_2 \rightarrow VX_c$$

$$T \rightarrow X_a Y_1$$

$$V \rightarrow X_a Y_2$$

Note that we can reuse Y_1, Y_2 for two productions