



Chapter 5

Parsing and recognition

- 5.1 Recognition and parsing in general grammars
- 5.2 Earley's method
- 5.3 Top-down parsing
- 5.4 Removing LL(1) conflicts
- 5.5 Bottom-up parsing

5.1 Recognition and parsing in general grammars

Cocke, Younger, and Kasami

$w \in L(G)$?

$w = a_1 a_2 \dots a_n$ and $w[i..j] = a_i \dots a_j$

A in $U[i, j]$ iff $A \Rightarrow^* w[i..j]$

initialization

A in $U[i, i]$ iff $A \rightarrow a_i$

recursion

A in $U[i, j]$ iff $A \rightarrow BC$ and

B in $U[i, k]$, C in $U[k + 1, j]$ ($i \leq k < j$)

Thm. $w \in L(G)$ for given CFG G in $\mathcal{O}(n^3)$ $n = |w|$

dynamic programming

A in $U[i, i]$ iff

$A \rightarrow a_i$

$w = cabab$

A in $U[i, j]$ iff

$A \rightarrow BC$ and

B in $U[i, k]$,

C in $U[k + 1, j]$

$(i \leq k < j)$

i/j	1	2	3	4	5
1	C	–	A	A	S, B
2		A	S, B	–	C
3			S, B	–	C
4				A	S, B
5					S, B

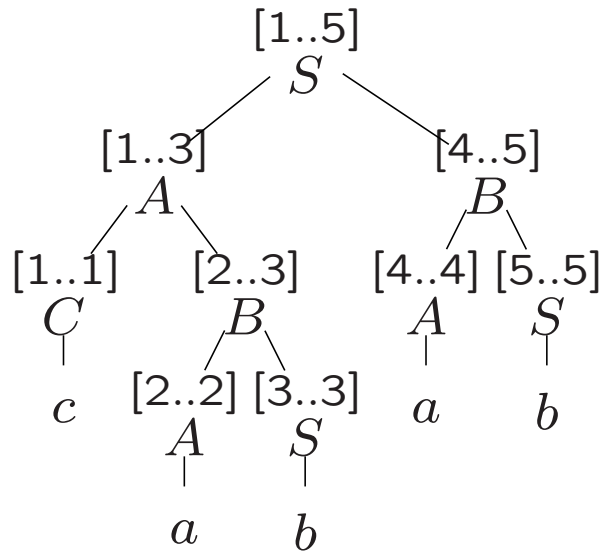
$S \rightarrow AB \mid b$

$A \rightarrow CB \mid AA \mid a$

$B \rightarrow AS \mid b$

$C \rightarrow BS \mid c$

$w \in L(G)$ as $S \in U[1, 5]$



- $S \rightarrow AB \mid b$
- $A \rightarrow CB \mid AA \mid a$
- $B \rightarrow AS \mid b$
- $C \rightarrow BS \mid c$

$w = cabab$

i/j	1	2	3	4	5
1	C	–	$A : (C, B, 1)$	$A : (A, A, 3)$	$S : (A, B, 3), (A, B, 4)$ $B : (A, S, 3), (A, S, 4)$
2		A	$S : (A, B, 2)$ $B : (A, S, 2)$	–	$C : (B, S, 3)$
3			S, B	–	$C : (B, S, 3)$
4				A	$S : (A, B, 4)$ $B : (A, S, 4)$
5					S, B

Thm. parse tree in $\mathcal{O}(n^3)$ steps, $n = |w|$

5.2 Earley's method

item $A \rightarrow \alpha \bullet \beta$ for $A \rightarrow \alpha \beta$

complete item $A \rightarrow \alpha \bullet$

(M_{ij})

$w = a_1 a_2 \dots a_n$

$A \rightarrow \alpha \bullet \beta$ in M_{ij}

$S \Rightarrow^* w[1..i]A\delta$ and $\alpha \Rightarrow^* w[i + 1, j]$

make-early-table

Predictor, Scanner, Completer (?)

- A. for every $S \rightarrow \gamma$ in P
put $S \rightarrow \bullet \gamma$ in M_{00}
- B. for $A \rightarrow \alpha \bullet a_{j+1} \beta$ in M_{ij}
put $A \rightarrow \alpha a_{j+1} \bullet \beta$ in M_{ij+1}
- C. if $A \rightarrow \alpha \bullet B \beta$ in M_{ij} and $B \rightarrow \gamma \bullet$ in M_{jk}
put $A \rightarrow \alpha a_{j+1} \bullet \beta$ in M_{ij+1}
- D. if $A \rightarrow \alpha \bullet B \beta$ in M_{ij} then
for every $B \rightarrow \gamma$ in P put $B \rightarrow \bullet \gamma$ in M_{jj}

Theorem 5.2.3

$w \in L$ iff $S \rightarrow \alpha \bullet \in M_{0,n}$ for some α

$$S \rightarrow T + S \mid T$$

$$T \rightarrow F * T \mid F$$

$$F \rightarrow (S) \mid a$$

$$M_{0,0} \left| \begin{array}{lll} S \rightarrow \bullet T + S & S \rightarrow \bullet T & T \rightarrow \bullet F * T \\ T \rightarrow \bullet F & F \rightarrow \bullet (S) & F \rightarrow \bullet a \end{array} \right.$$

$$w = (a + a) * a$$

$$\begin{array}{l} M_{0,1} \\ M_{0,2} \\ \dots \end{array} \left| \begin{array}{l} F \rightarrow (\bullet S) \\ F \rightarrow (S \bullet) \\ \dots \end{array} \right.$$

5.3 Top-down parsing

expressions, terms

$$\Sigma = \{ a, +, [,] \}$$

$$V = \{ E, Z, X, T \}$$

$$E \rightarrow TZ$$

$$Z \rightarrow X \mid \epsilon$$

$$X \rightarrow +TZ$$

$$T \rightarrow a \mid [E]$$

	E	Z	X	T
a	$E \rightarrow TZ$			$T \rightarrow a$
$+$		$Z \rightarrow X$	$X \rightarrow +TZ$	
$[$	$E \rightarrow TZ$			$T \rightarrow [E]$
$]$		$Z \rightarrow \epsilon$		
$\$$		$Z \rightarrow \epsilon$		

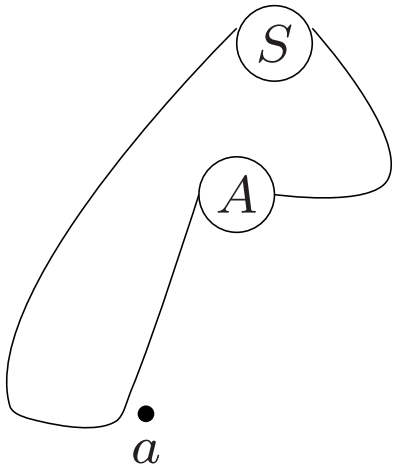
$$z = a + [a + a]$$

$$E \xRightarrow{E,a} TZ \xRightarrow{T,a} aX \xRightarrow{X,+} a+TZ \xRightarrow{T,[}$$

$$a+[E]Z \xRightarrow{E,a} a+[TZ]Z \xRightarrow{T,a} a+[aZ]Z \xRightarrow{Z,+}$$

$$a+[aX]Z \xRightarrow{X,+} a+[a+TZ]Z \xRightarrow{T,a} a+[a+aZ]Z \xRightarrow{Z,]}$$

$$a+[a+a]Z \xRightarrow{Z,\$} a+[a+aZ]$$



$$S \Rightarrow_{\ell}^* uAv \Rightarrow_{\ell}^* uax = z$$

$$(u \in \Sigma^*, a \in \Sigma, A \in V, v \in \Sigma^*)$$

A and a determine production $A \rightarrow \alpha$

$$\text{first}(\alpha) = \{x \in \Sigma^* \mid \alpha \Rightarrow^* x\}$$

$$\text{follow}(A) = \{x \in \Sigma^* \mid S \Rightarrow_{\ell}^* uAv \text{ and } v \Rightarrow^* x\}$$

$$a \in \text{init}_1(\text{first}(\alpha) \text{ follow}(A)) \text{ for some } A \rightarrow \alpha$$

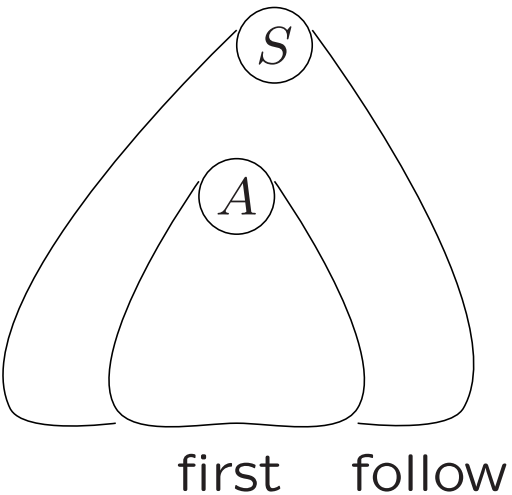
LL(k): first k symbols

here $k = 1$

complication when ϵ in first

$$\text{FIRST}(u) = \text{init}_1(\text{first}(u))$$

$$\text{FOLLOW}(A) = \text{init}_1(\text{follow}(A))$$



$$\text{init}_1(\epsilon) = \epsilon, \text{init}_1(ax) = a$$

$$E \rightarrow TZ$$

$$Z \rightarrow X \mid \epsilon$$

$$X \rightarrow +TZ$$

$$T \rightarrow a \mid [E]$$

i	E	Z	X	T
0	\emptyset	ϵ	$+$	$a, [$
1	$a, [$	$\epsilon, +$	$+$	$a, [$
2	$a, [$	$\epsilon, +$	$+$	$a, [$

$$F_i(a) = \{a\}$$

$$F_0(A) = \{ a \in \Sigma \mid A \rightarrow ay \in P \} \\ \cup \{ \epsilon \mid A \rightarrow \epsilon \in P \}$$

$$F_{i+1}(A) = F_i(A) \cup$$

$$\bigcup_{A \rightarrow Y_1 \dots Y_m \in P} \text{init}_1(F_i(Y_1) \cdot \dots \cdot F_i(Y_m))$$

repeat until $F_{i+1} = F_i$

Thm. $a \in \text{FIRST}(A)$ iff $a \in \bigcup_{i \in \mathbb{N}} F_i(A)$

Prf. if $A \Rightarrow^n ax$ ($a \in \Sigma$) then $a \in F_n(A)$

$$\text{FIRST}(A) = \text{init}_1\{x \in \Sigma^* \mid A \Rightarrow^* x\}$$

if $A \Rightarrow^n ax$ ($a \in \Sigma$) then $a \in F_n(A)$

$n = 1$, $A \Rightarrow ax$, hence $a = \text{init}_1(F_0(a)F_0(x))$ for the production $A \rightarrow ax$

may be empty?

$$E \rightarrow TZ$$

$$Z \rightarrow X \mid \epsilon$$

$$X \rightarrow +TZ$$

$$T \rightarrow a \mid [E]$$

i	E	Z	X	T
0], \$	\emptyset	\emptyset	+
1], \$], \$	\emptyset	+,], \$
2], \$], \$], \$	+,], \$
3], \$], \$], \$	+,], \$

$$FL_0(A) =$$

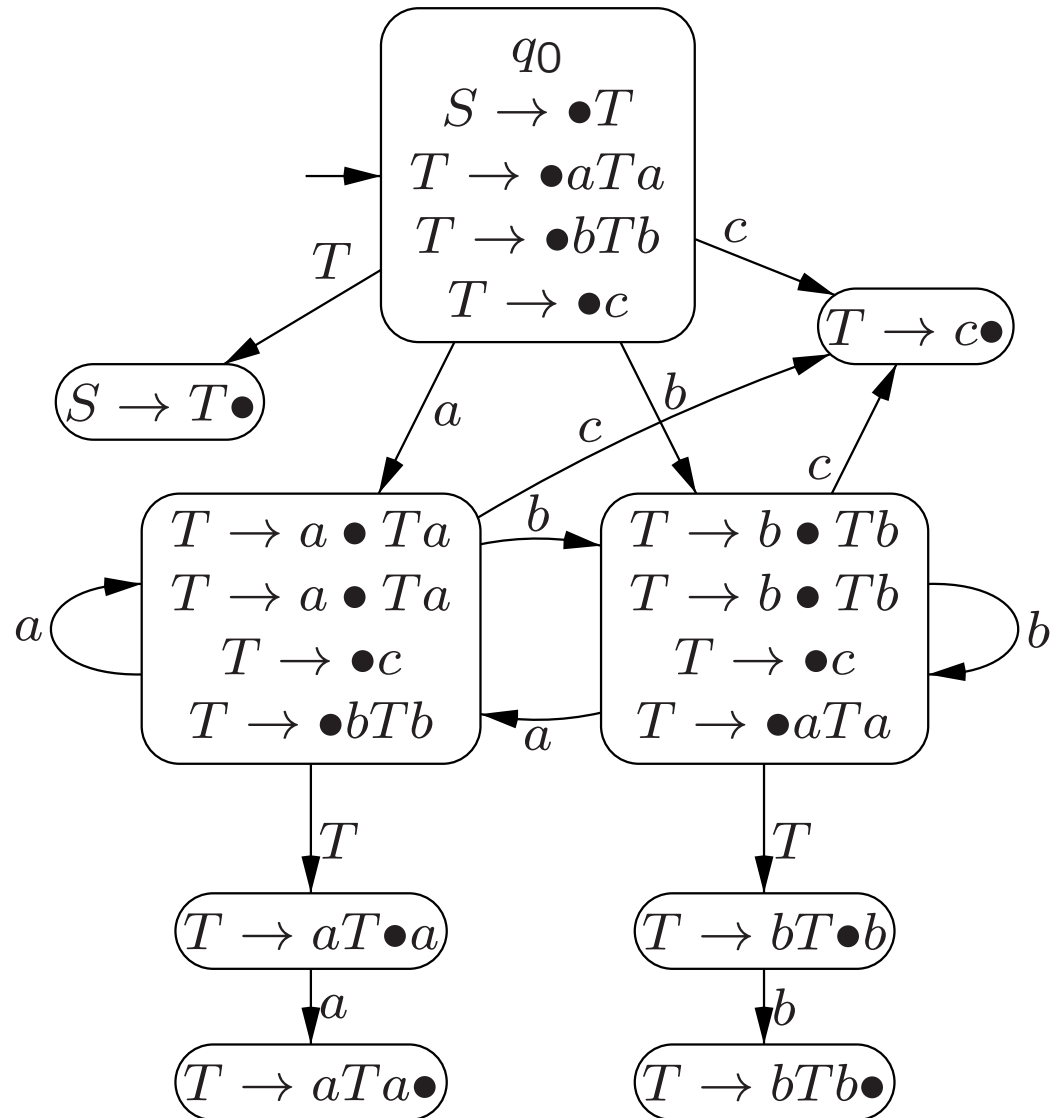
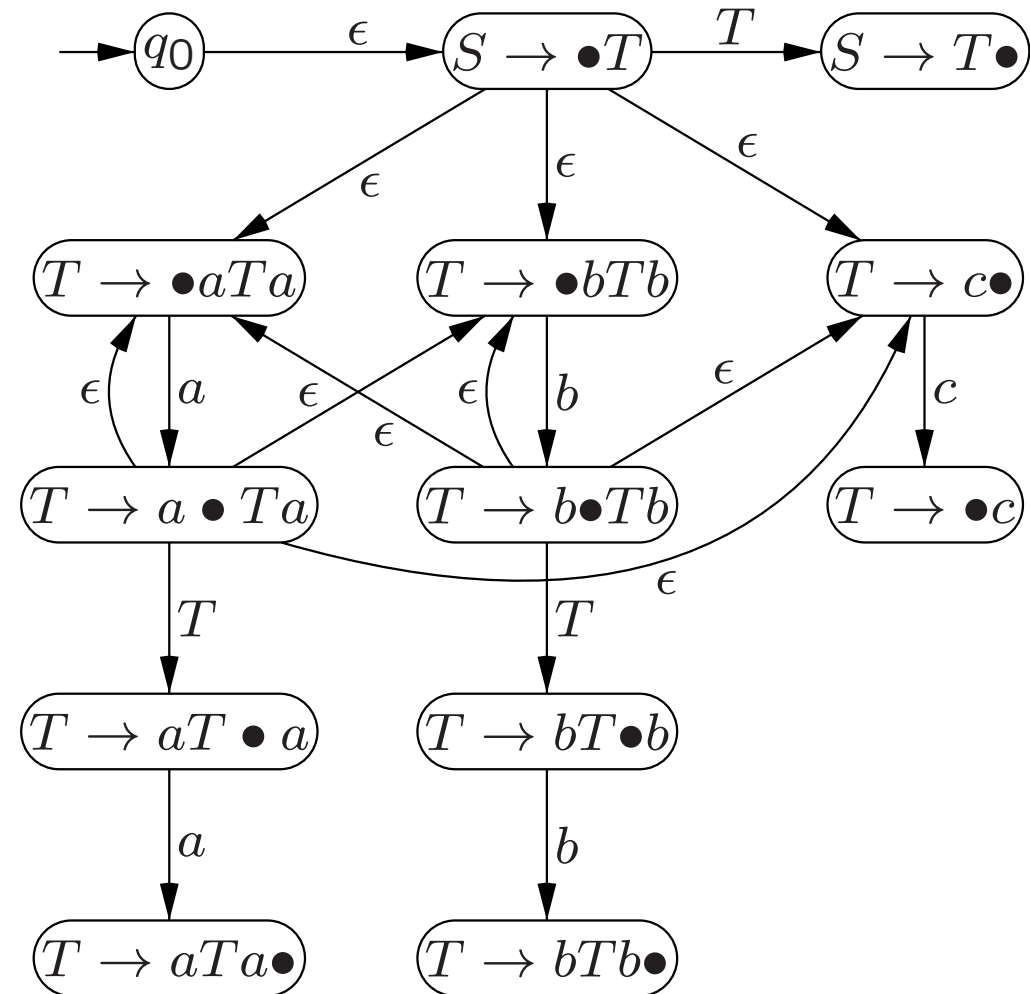
$$\{ a \in \Sigma \mid B \rightarrow uAv, a \in \text{FIRST}(v) \} \cup \{ \$ \mid A = S \}$$

$$FL_{i+1}(A) = FL_i(A) \cup$$

$$\bigcup_{\substack{B \rightarrow uAv \in P \\ \epsilon \in \text{FIRST}(v)}} FL_i(B)$$

Thm. $a \in \text{FOLLOW}(A)$ iff $a \in \bigcup_{i \in \mathbb{N}} FL_i(A)$

5.5 Bottom-up Parsing



transparencies made for

Second Course in
Formal Languages and
Automata Theory

based on the book by Jeffrey Shallit
of the same title

Hendrik Jan Hoogeboom, Leiden
<http://www.liacs.nl/~hoogeboo/second/>