Chapter 5

Parsing and recognition

5.1 Recognition and parsing in general grammars

Cocke, Younger, and Kasami

$w \in L(G)$?

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

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

initialization

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

recursion

A in $U[i,j]$ iff $A \to BC$ and
B in $U[i,k]$, C in $U[k+1,j]$ $(i \le k < j)$

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

dynamic programming

A in $U[i,i]$ iff

$A \to a_i$

A in $U[i,j]$ iff

$A \to BC$ and

B in $U[i,k]$,

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

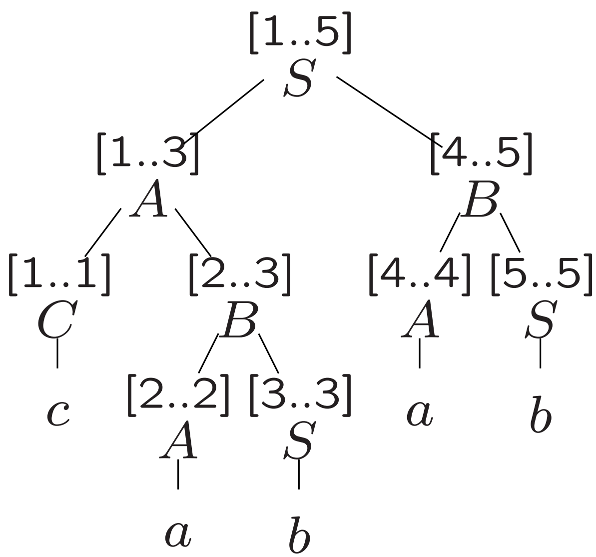
$(i \le k < j)$

$$\begin{array}{lcl} S & \to & AB \mid b \\ A & \to & CB \mid AA \mid a \\ B & \to & AS \mid b \\ C & \to & BS \mid c \end{array}$$

$w = cabab$

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

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



$$\begin{aligned} S &\rightarrow AB \mid b \\ A &\rightarrow CB \mid AA \mid a \\ B &\rightarrow AS \mid b \\ C &\rightarrow BS \mid c \end{aligned}$$

$w = cabab$

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

V 5 *Earley's method*

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

complete item $A \to \alpha \bullet$

(M_{ij})

$w = a_1 a_2 \ldots a_n$

$A \to \alpha \bullet \beta$ 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 $S \to \gamma$ in P
put $S \to \bullet\gamma$ in M_{00}

B. for $A \to \alpha \bullet a_{j+1}\beta$ in M_{ij}
put $A \to \alpha a_{j+1} \bullet \beta$ in M_{ij+1}

C. if $A \to \alpha \bullet B\beta$ in M_{ij} and $B \to \gamma\bullet$ in M_{jk}
put $A \to \alpha a_{j+1} \bullet \beta$ in M_{ij+1}

D. if $A \to \alpha \bullet B\beta$ in M_{ij} then
for every $B \to \gamma$ in P put $B \to \bullet\gamma$ in M_{jj}

Theorem 5.2.3

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

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

$$\begin{array}{c|lll} M0,0 & S \to \bullet T + S & S \to \bullet T & T \to \bullet F * T \\ & T \to \bullet F & F \to \bullet(S) & F \to \bullet a \end{array}$$

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

$$\begin{array}{l|l} M_{0,1} & F \to (\bullet S) \\ M_{0,2} & F \to (S \bullet) \\ \ldots & \ldots \end{array}$$

5.3 Top-down parsing

expressions, terms

$\Sigma = \{\, a, +, [,] \,\}$
$V = \{\, E, Z, X, T \,\}$

$E \to TZ$

$Z \to X \mid \epsilon$

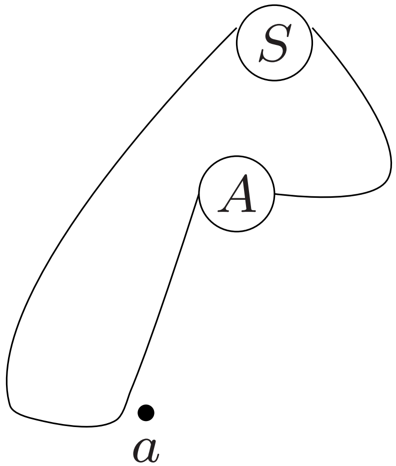
$X \to +TZ$

$T \to a \mid [E]$

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

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

$$E \overset{E,a}{\Rightarrow} TZ \overset{T,a}{\Rightarrow} aX \overset{X,+}{\Rightarrow} a{+}TZ \overset{T,[}{\Rightarrow}$$
$$a{+}[E]Z \overset{E,a}{\Rightarrow} a{+}[TZ]Z \overset{T,a}{\Rightarrow} a{+}[aZ]Z \overset{Z,+}{\Rightarrow}$$
$$a{+}[aX]Z \overset{X,+}{\Rightarrow} a{+}[a+TZ]Z \overset{T,a}{\Rightarrow} a{+}[a+aZ]Z \overset{Z,]}{\Rightarrow}$$
$$a{+}[a+a]Z \overset{Z,\$}{\Rightarrow} a{+}[a+aZ]$$



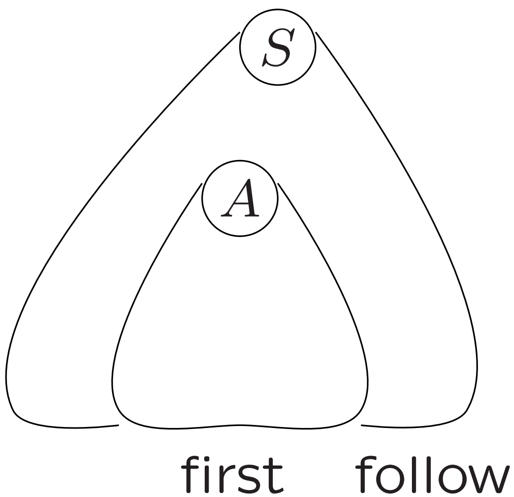
$$S \Rightarrow_\ell^* uAv \Rightarrow_\ell^* uax = z$$
$$(u \in \Sigma^*,\ a \in \Sigma,\ A \in V,\ v \in \Sigma^*)$$

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

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

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

$a \in$ init$_1$(first(α) follow(A)) for some $A \to \alpha$



LL(k): first k symbols here $k = 1$

complication when ϵ in first

FIRST$(u) =$ init$_1$(first(u))

FOLLOW$(A) =$ init$_1$(follow(A))

init$_1(\epsilon) = \epsilon$, init$_1(ax) = a$

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

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

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

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

$$F_{i+1}(A) = F_i(A) \cup \bigcup_{A \to Y_1 \cdots Y_m \in P} \mathrm{init}_1(F_i(Y_1) \cdot \ldots \cdot F_i(Y_m))$$

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

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

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

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

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

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

may be empty?

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

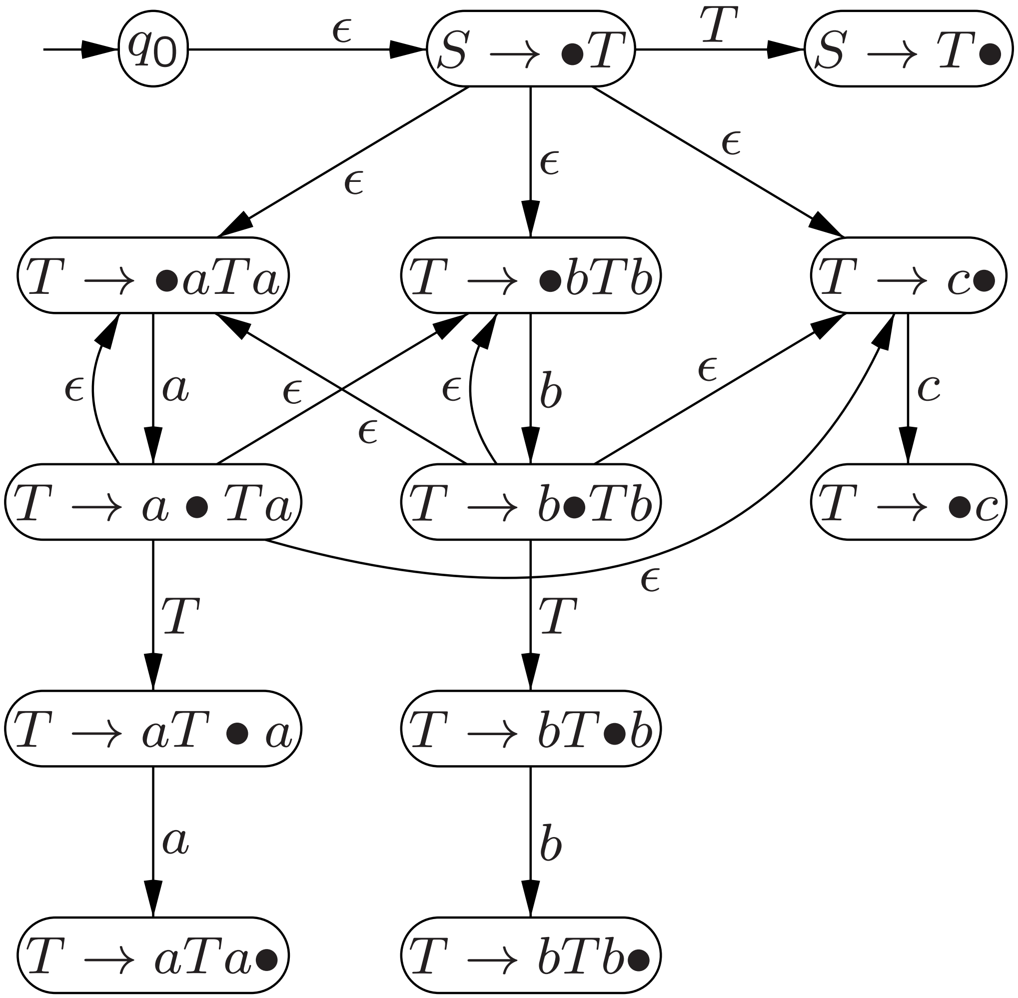
i	E	Z	X	T
0	], \$	$\varnothing$	$\varnothing$	+
1	], \$	], \$	$\varnothing$	+,], \$
2	], \$	], \$	], \$	+,], \$
3	], \$	], \$	], \$	+,], \$

$$FL_0(A) = \{\, a \in \Sigma \mid B \to uAv, a \in \mathsf{FIRST}(v) \,\} \cup \{\, \$ \mid A = S \,\}$$

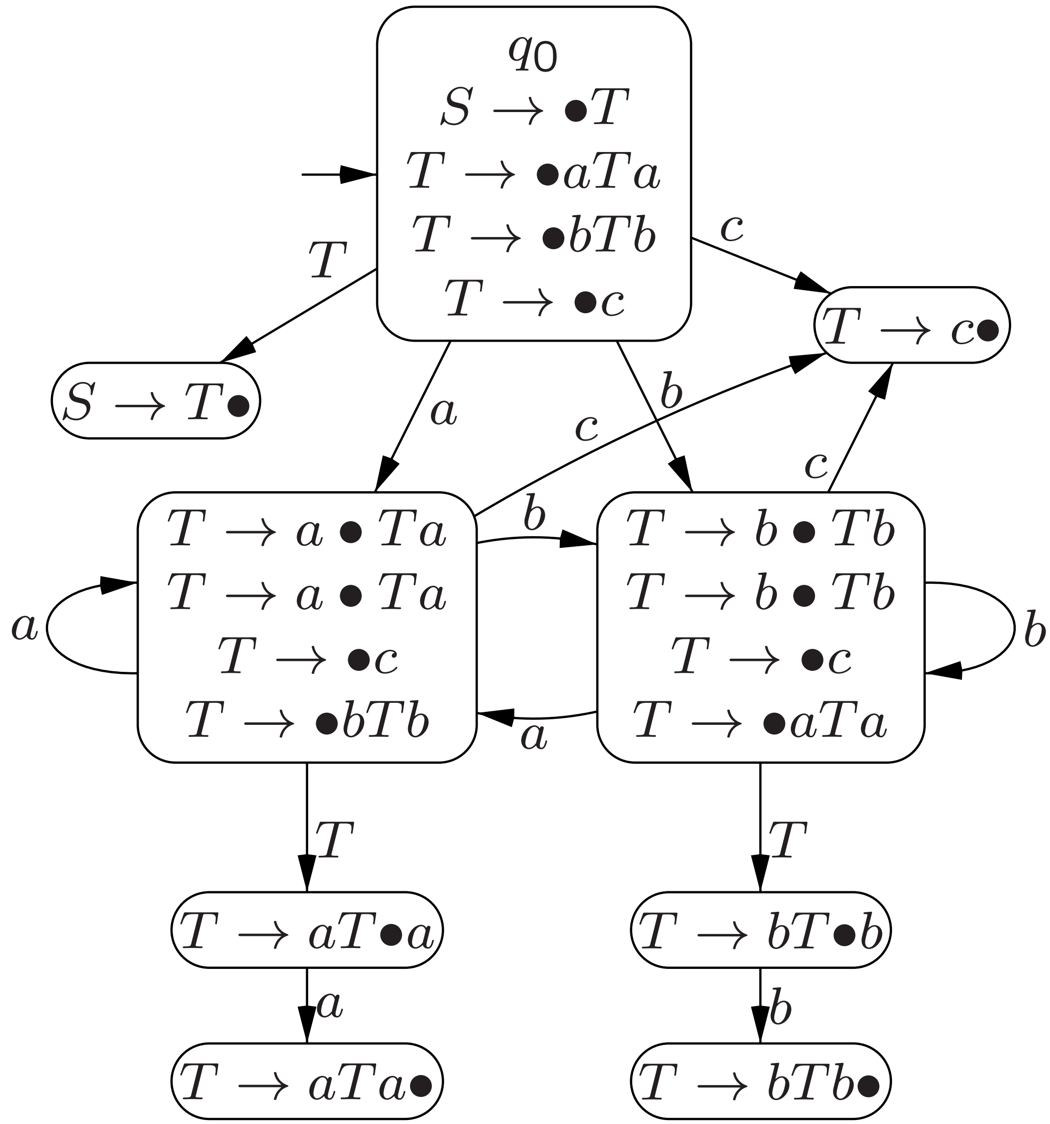
$$FL_{i+1}(A) = FL_i(A) \cup \bigcup_{\substack{B \to uAv \in P \\ \epsilon \in \mathsf{FIRST}(v)}} FL_i(B)$$

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

5.5 Bottom-up Parsing



q_0
$S \to \bullet T$
$S \to T\bullet$
$T \to \bullet aTa$
$T \to \bullet bTb$
$T \to c\bullet$
$T \to a \bullet Ta$
$T \to b\bullet Tb$
$T \to \bullet c$
$T \to aT \bullet a$
$T \to bT\bullet b$
$T \to aTa\bullet$
$T \to bTb\bullet$
ϵ
T
a
b
c



q_0
$S \to \bullet T$
$T \to \bullet aTa$
$T \to \bullet bTb$
$T \to \bullet c$
$T \to c\bullet$
$S \to T\bullet$
$T \to a \bullet Ta$
$T \to a \bullet Ta$
$T \to \bullet c$
$T \to \bullet bTb$
$T \to b \bullet Tb$
$T \to b \bullet Tb$
$T \to \bullet c$
$T \to \bullet aTa$
$T \to aT\bullet a$
$T \to bT\bullet b$
$T \to aTa\bullet$
$T \to bTb\bullet$
T
a
b
c

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/`