COL 352 Lecture 18, 19 Feb 20, 21

Commonly used CFL

Arithmetic expressions

$$S \rightarrow S+S \mid S-S \mid S\times S \mid S\div S \mid (S) \mid V \mid N$$

$$V \rightarrow \{a,b,c \ldots z\}\cdot V \mid \{a,b,c \ldots z\}$$ variable

$$N \rightarrow \{0,1,2\ldots 9\}\cdot N \mid \{0,1,2\ldots 9\}$$ number

that produces expression like

$$(x+9)\,4y + 10 \quad \text{etc.}$$

Balanced parenthesis

$$S \rightarrow (S) \mid S\cdot S \mid ()$$

that produces $(())\;()\;((()))$ etc.

Conversion of Arbitrary CFG to normal forms

Chomsky Normal Form	Greibach Normal Form
Prodn of the form $A \to BC \mid a$ — exactly 2 variables or one terminal on the Right hand side	$A \to a\alpha$, $\alpha \in \{T \cup V\}^*$ — i.e. it must start with a terminal

Steps of transformation

(i) Eliminate useless symbols:

If a variable X doesn't appear in any derivation, i.e. $S \xrightarrow{*} \alpha_1 \cancel{X} \alpha_2$

or, $X \not\xrightarrow{*} T^*$ X doesn't lead to any string over terminals, such variables (and all rules containing them) can be discarded without changing the language

(ii) Eliminate ϵ prodns.

Suppose $S \to AB\alpha$ and $A \to \epsilon$, $B \to \epsilon$

then we can eliminate $A \to \epsilon$, $B \to \epsilon$ and add the following rules in lieu of them

$$S \to A\alpha \mid S \to B\alpha \mid S \to AB\alpha \mid S \to \alpha$$

etc. and all derivations of the original grammar can be preserved

(iii) Unit prodn $A \to B$ $B \to C$ ... can also be eliminated like

$$A \cancel{\to} B \qquad B \to \alpha \qquad \text{then } A \to \alpha$$

<u>Thm</u>: Given any arbitrary CFG $\mathcal{G}$ it can be transformed to G_1, G_2 where G_1 is in CNF and G_2 in GNF s.t.

$$L(\mathcal{G}) = L(G_1) = L(G_2)$$

We will show several applications of the normal forms as they are much easier to work with and have nice properties.

. .

CNF $A \to BC \mid a$

$S \to aB \mid bA$ $\qquad$ $S \to C_a B \mid C_b A$

$A \to a \mid aS \mid bAA$ $\qquad$ $A \to a \mid C_a S \mid C_b D$

$B \to b \mid bS \mid aBB$ $\qquad$ $B \to b \mid C_b S \mid C_a E$

$C_a \to a \quad C_b \to b$

$D \to AA$

$E \to BB$

Membership problem Given $G = (V, T, S, P)$ in CNF and a string $w \in \Sigma^*$ does $S \xrightarrow{*} w$

$|w| = n \qquad w_1 w_2 w_3 \ldots w_n$

$w_i \in T$

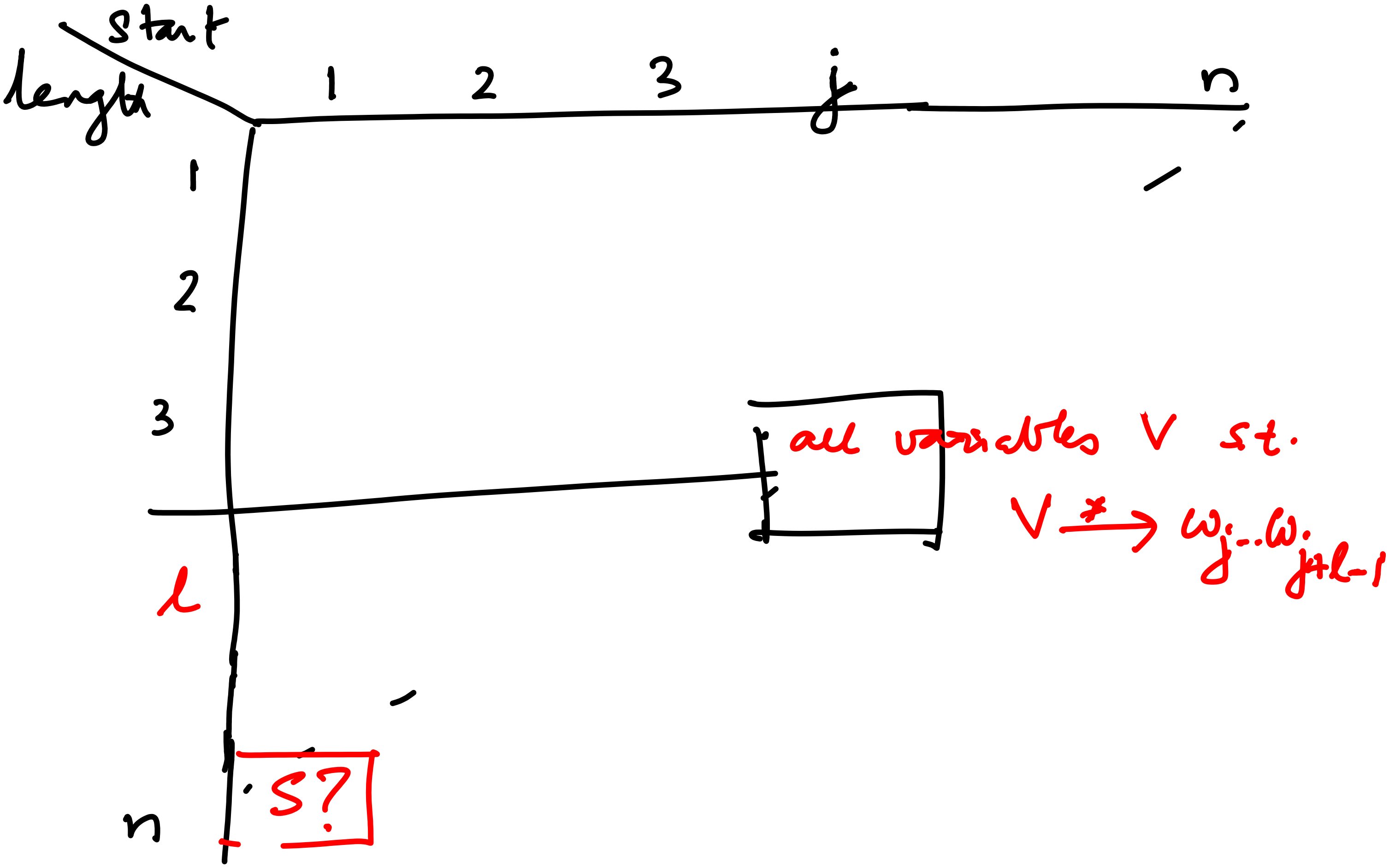
$S \overset{?}{\xrightarrow{*}} w$

$S \xrightarrow{*} w$ iff there is a j s.t.

$S \to AB$ and $A \xrightarrow{*} w_1 w_2 \ldots w_j$

and $B \xrightarrow{*} w_{j+1} w_{j+2} \ldots w_n$

$w_{ij} = w_i w_{i+1} \ldots w_j$



$S \to C_a B \mid C_b A$

$A \to a \mid C_a S \mid C_b D$

$B \to b \mid C_b S \mid C_a E$

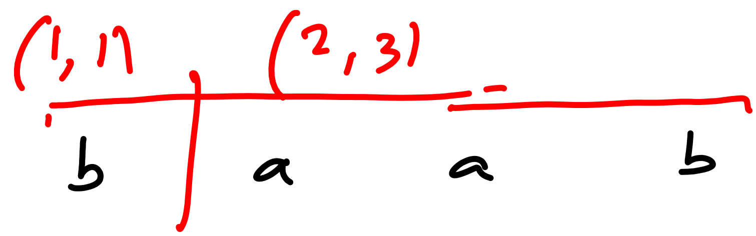
$D \to A A$

$E \to B B$

$C_a \to a$

$C_b \to b$

Ex. baab



length \ start	1	2	3	4
1	B, C_b	A, C_a	A, C_a	B, C_b
2	S	D	S	
3	A	A		
4	S			

$S \to C_b A$.

$C_b \to b \quad A \to aab$

The D.P. takes $O(n^3)$ steps (considering the size of grammar to be constant and ignoring data structure cost)

CYK algorithm