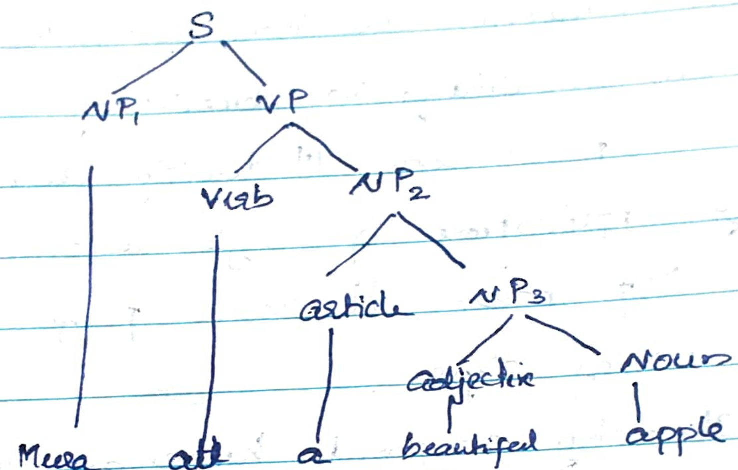
Grammar

Grammar is a set of rules to define a particular language meaningfully

eg: Meera ate a beautiful apple.

It's a sentence.



This tree is called syntax tree or parse tree or derivation tree

$S \rightarrow \langle NP_1 \rangle \langle VP \rangle$

$NP_1 \rightarrow$ meera

$VP \rightarrow \langle verb \rangle \langle NP_2 \rangle$

$VP \rightarrow$ ate

$NP_2 \rightarrow \langle article \rangle \langle NP_3 \rangle$

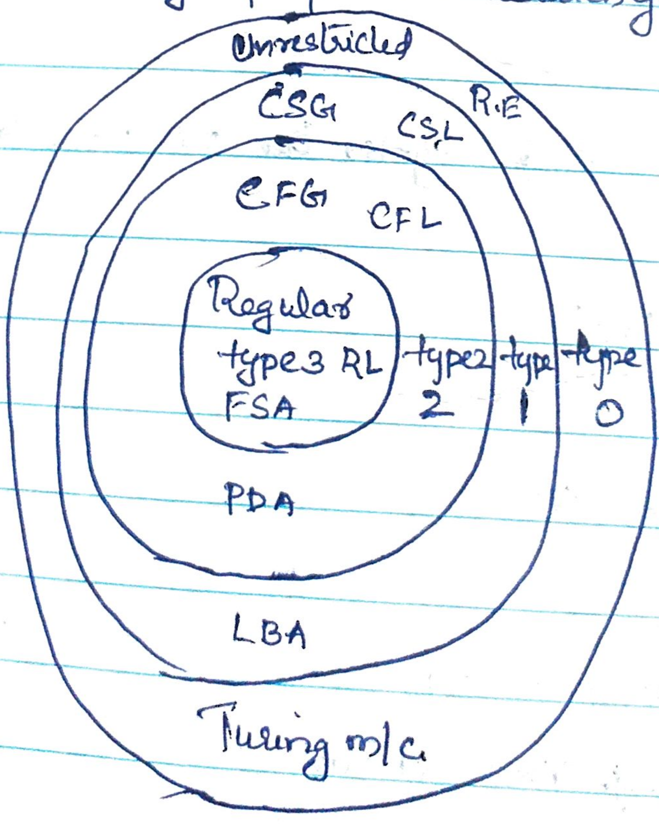
$NP_3 \rightarrow \langle adjective \rangle \langle noun \rangle$

Noun $\rightarrow$ apple

$\langle adjective \rangle \rightarrow$ beautiful.

These are called set of production rules.
At one time, only one rule is applied and replaced by terminals or non-terminals

* Terminals are those symbols which derivation terminates. The terminals can't be further divided
* Non-terminals are those which can be derived from something to something
* $\alpha \Rightarrow \beta$ means it directly derives β from α by application of single rule
* $\alpha \overset{*}{\Rightarrow} \beta$ means it derives β from α through a no: of steps.
* Chomsky proposed hierarchy of grammars



Type 0	Grammar	-	Unrestricted Grammar
	Language	-	Recursively
	Automaton	-	Turing m/c
Type 1	Grammar	-	Context sensitive
	Language	-	Context sensitive
	Automation	-	Linear Bounded
Type 2	Grammar	-	Context - Free
	Language	-	Context - Free
	Automation	-	Pushdown
Type 3	Grammar	-	Regular
	Language	-	Regular
	Automaton	-	Finite Automata

Type 0: $\alpha\beta\gamma \to A\beta B$ L.H.S = $(V \cup T)^* \, V \, (V \cup T)^*$

R.H.S = $(V \cup T)^* \, V \, (V \cup T)^*$

Type 1: $|\alpha\beta\gamma| \to |\alpha\beta|$ $|LHS| \leq |RHS|$

Type 2: $V \to (V \cup T)^*$ L.H.S contains only one variable.

Type 3: $V \to aB$ R.H.S starts a terminal foll. by a variable

$V \to a$ Contains only one terminal.

Type 0

$x \to y$ $\quad x \in (V \cup T)^* V (V \cup T)^*$

$y \in (V \cup T)^*$

LHS contains atleast one variable

Type 1 (CSG)

$x \to y$ $\quad \alpha A \beta \to \alpha \gamma \beta$ $\quad$ [$S \to \epsilon$, but not in

$|x| \leq |y|$ $\quad x \in (V \cup T)^* V (V \cup T)^*$ $\quad$ S, RHS]

$y \in (V \cup T)^*$

A can be replaced by γ in the context of α in left & β is right to A.

Type 2 - Context free grammar. Free to context.

$x \to y$ $\quad x \in V$ - exactly one variable.

$y \in (V \cup T)^*$

y can be replaced by x in any context.

Type 3

Regular Grammar is right linear or left linear. but not both

$A \to aB \mid Ba$

$A \to a$

* Grammar can be represented by 4 tuples $G = (V, T, P, S)$

where V - Set of variables or nonterminals can be represented by Capital letters A, B ... Z denoted V_n

T - Set of terminals can't be further divided represented with small letters a, b, ... z denoted by V_t

P - Set of pdns that represent the recursive defn of the language.

S - Start symbol.

Derivation

* If $\alpha \rightarrow \beta$ is a pdn ~~in~~ P in CFG and a & b are strings in $(V_n \cup V_t)^*$, then $a\alpha b \underset{G}{\Rightarrow} a\beta b$, β directly derives from α.

* If $\alpha_1 \Rightarrow \alpha_2$, $\alpha_2 \Rightarrow \alpha_3$, $\alpha_3 \Rightarrow \alpha_4$ ....

$\alpha_1 \overset{*}{\Rightarrow} \alpha_4$ α_4 multiply derives from α_1

* Derivation from start symbol to terminal based on pdn rule and at each step only one rule is applied at one time. That tree is called derivation tree or parse tree or syntax tree

* Derive something from something is derivation

2 types of Derivations

1) Leftmost derivation · leftmost symbol is replaced at each step

2) Rightmost derivation - Rightmost symbol is replaced at each step.

It's not necessary that, it always do left most or rightmost derivation to get terminals. Any variable can be replaced in any order and get the terminal string

eg: $S \to A1B$

$A \to 0A \mid \epsilon$

$B \to 0B \mid 1B \mid \epsilon$

Generate 00101 using leftmost & Rightmost derivation

1) LM derivation

$S \to \underline{A}1B$

$S \to 0\underline{A}1B \quad (A \to 0A)$

$S \to 00\underline{A}1B \quad (A \to 0A)$

$S \to 00\epsilon 1B \quad (A \to \epsilon)$

$S \to 001\underline{B}$

$S \to 0010\underline{B} \quad (B \to 0B)$

$S \to 00101B \quad (B \to 1B)$

$S \to 00101\epsilon \quad (B \to \epsilon)$

2. **RM derivation**

$S \to A1\underline{B}$

$S \to A10B$ $(B \to 0B)$

$S \to A10\underline{B}$ $(B \to 1B)$

$S \to A101\underline{B}$ $(B \to \epsilon)$

$S \to A101$ $(A \to 0A)$

$S \to 0\underline{A}101$ $(A \to 0A)$

$S \to 00A101$ $(A \to \epsilon)$

$S \to 00101$

Q Generate a) 1001 b) 00011 using lm & Rm

Language of CFG

If $G = (V, T, P, S)$, the language of G denoted by $L(G)$ is a set of terminals that have derivation from the start symbol.

$L(G) = \{ w \text{ in } T \mid S \overset{*}{\underset{G}{\Rightarrow}} w \}$

Sentential Form

String which in the form of $(V \cup T)^*$ from something to something tho' the production rule at one time

$S \Rightarrow \alpha$, α is a sentential form eg1: Consider $G = (V, t, p, s)$, $V = \{s\}$, $t = \{a, b\}$

$S \rightarrow aSb$ — (1)

$S \rightarrow ab$ — (2)

Find out the language generated

A: $S \rightarrow aSb$

$S \rightarrow aaSbb$ $(\because S \rightarrow aSb)$

$S \rightarrow aaaSbbb$

$S \rightarrow aaaabbbb$ $(S \rightarrow ab)$

$L(G) = a^n b^n \mid n \geq 1$

Q2: Design a CFG for the foll. language

$L = \{ wcw^R \mid w \in (a,b)^* \}$

A $G = (V, t, P, S)$

$V = \{S\}$

$t = \{a, b\}$

A $S \rightarrow aSa \mid bSb \mid c$

$S \rightarrow aSa$

$S \rightarrow aaSaa$

$S \rightarrow aabSbaa$

Q2. Write a CFG. which generates string of balanced parenthesis

A $G = (V, t, P, S)$, $V = \{S\}$, $t = \{(,)\}$

$S \rightarrow SS$

$S \rightarrow (S)$

$S \rightarrow \epsilon$

((())), ()() both will be accepted

Q3: Design a CFG which generates palindrome for binary numbers.

$G = (V, t, P, S)$, $t = \{0, 1\}$

$S \rightarrow 0S0 \mid 1S1$

$S \rightarrow 0 \mid 1 \mid \epsilon$

$S \rightarrow 0S0$

$S \rightarrow 01S10$

$S \rightarrow 01010$

Q4: Write a CFG for R.E $r = 0^* \mid (0+1)^*$

$G = (V, t, P, S)$, $t = \{0, 1\}$, $V = \{S, A, B\}$.

$S \rightarrow A \mid B$

$A \rightarrow 0A \mid \epsilon$

$B \rightarrow 0B \mid 1B \mid \epsilon$

Q5. Write a CFG which generates strings having equal no of a's & b's

G = (V, T, P, S)

T = {a, b}

V = {S}

S = {S}.

$S \rightarrow aSbS \mid bSaS \mid \epsilon$

A Q6: Design CFG having any combination of a's and b's except null string. G = (V, T, P, S), V = {S}

T = {a, b}

A: G = (V, T, P, S), V = {S}, T = {a, b}.

$S \rightarrow aS$

$S \rightarrow bS$

$S \rightarrow a \mid b$

A Q7: Design CFG for r.e $r = (a+b)^* aa (a+b)^*$

G = (V, T, P, S)

V = {S, A}

T = {a, b}

$S \rightarrow AaaA$

$A \rightarrow aA \mid bA \mid \epsilon$

Q8: Write CFG for ww^r $\Sigma = \{0,1\}$

$S \rightarrow 0S0$

$S \rightarrow 1S1$

$S \rightarrow \epsilon$

Q9: $L(G) = \{ab\,(bbaa)^n\, bba\, (ba)^n : n \geq 0\}$

$S \rightarrow abX$

$X \rightarrow bbaaXba$

$X \rightarrow bba$

Q10: Same Qn where $n \geq 0$.

$S \rightarrow abX$

$X \rightarrow bbYa$

$Y \rightarrow \epsilon$

$Y \rightarrow aaXb$.

Q11: Design CFG for $L = \{ a^n b^m : n \neq m \}$

2 cases $n > m$ $G = \{v, t, p, s\}$

$N = \{S_2, A, S\}$

$t = \{a, b\}$

$n > m$

$S_{21} \to AS_1$

$A \to aA \mid a$

$S_1 \to aS_1b \mid \epsilon$

$n < m$

$S_2 \to S_1B$

$S_1 \to aS_1b \mid \epsilon$

$B \to bB \mid b$

$S \to S_{21} \mid S_2$

A Q12: Design CFG for $L = \{ 0^n 1^n \mid n \geq 0 \cup 1^n 0^n \mid n \geq 0 \}$

$S \to S_1 \mid S_2$

$S_1 \to 0S_11 \mid \epsilon$

$S_2 \to 1S_20 \mid \epsilon$

A Q13: Design CFG for $\Sigma = \{a,b\}$

a) all strings with exactly one a

b) all strings with atleast one a

c) all strings with atleast 3 a's

a) $G = (V, T, P, S)$ $V = \{S, A\}$, $T = \{a, b\}$

$S \rightarrow AaA$

$A \rightarrow bA \mid \epsilon$

b) $S \rightarrow AaA$

$A \rightarrow aA \mid bA \mid \epsilon$

c) $G = (V, T, P, S)$, $V = \{S, A\}$, $T = \{a, b\}$.

$S \rightarrow AaAaAaA$

$A \rightarrow aA \mid bA \mid \epsilon$

Q 14: Design CFG for the $L = \{a^n b^n c^m d^m \mid n \geq 1, m \geq 1\}$.

$S \rightarrow XY$

$X \rightarrow aXb \mid ab$

$Y \rightarrow cYd \mid cd$

$G = (V, T, P, S)$

$V = \{S, X, Y\}$.

$T = \{a, b, c, d\}$.

A Q13. Design CFG for $\{a^n b^m c^m d^n \mid n \geq 1, m \geq 1\}$

$S \to aSd$

$S \to aAd$

$A \to bAc \mid bc$.

A Q14: Design CFG for $L = \{a^n b^{2n} \mid n \geq 0\}$.

$S \to aSbb \mid \epsilon$

Q15: Design CFG for $L = \{a^{2n} b^m \mid n > 0, m \geq 0\}$

$S \to aaAB$

$A \to aaA \mid \epsilon$

$B \to bB \mid \epsilon$

Reduction of Context Free Grammar

1 Eliminating useless symbols

A Symbol in CFG is said to be useful iff

a) $y \overset{*}{\Rightarrow} \omega$ where $\omega \in L(G)$ and ω is V_t^*, y leads to a string of terminals, y is said to be generating

b) If there is a derivation $S \overset{*}{\Rightarrow} \alpha y \beta \overset{*}{\Rightarrow} \omega$, $\omega \in L(G)$ then y is said to be reachable.

- A symbol is useful if both be generating and reachable

Q: Eliminate non-generating symbols and non-reachable symbol.

$S \rightarrow AB | a$

$A \rightarrow b$

$S \rightarrow AB$

$A \rightarrow b$ ✓

$A \rightarrow b$

$S \rightarrow a$ ✓

$S \rightarrow a$

A: $S \rightarrow AB$ is non-generating since $S \rightarrow bB$ which doesn't generate any terminals & A is non-reachable

$S \rightarrow a$.

Q2: Eliminate the useless symbol.

$S \to aB \mid bX$

$A \to BAd \mid bSX \mid a$

$B \to aSB \mid bBX$

$X \to SBD \mid aBX \mid ad$

A

$S \to bX$

$X \to ad$

Q3:

$A \to xyz \mid Xyzz$

$X \to Xz \mid xYz$

$Y \to yYy \mid Xz$

$Z \to Zy \mid z$

A:

$A \to xyz$

$Z \to Zy \mid z$

final : $A \to xyz$.

Q4:

$S \to aC \mid SB$ $\quad$ $C \to ad \mid aBC$.

$A \to bSCa$

$B \to aSB \mid bBC$

Q7: Construct a reduced grammar P: $S \to AB$

$A \to a$ $\quad E \to c$

$B \to b$ $\quad B \to C$

Ans: $S \to AB$

$A \to a$

$B \to b$

Q8: Construct a reduced grammar whose pds are

$S \to AB \mid CA$

$B \to BC \mid AB$

$A \to a$

$C \to aB \mid b$

Ans: $S \to CA$

$A \to a$

$C \to b$

Q9: Construct a reduced grammar equi to the grammar

$S \to aAa$, $A \to Sb \mid bCC \mid DaA$

$C \to abb \mid DD$, $E \to aC$ $\quad D \to aDA$

A: $S \to aC$

~~$A \to bSCa$~~

$C \to ad$

Q5: $P = \{ S \to A11B \mid 11A$

$S \to B11$

$A \to 0$

$B \to BB \}$.

A: $S \to 11A$

$A \to 0$

Q6: Find CFG with no useless symbols:

$S \to AB \mid CA$

$A \to a$

$B \to BC \mid AB$

$C \to aB \mid b$

A: $S \to CA$

$A \to a$

$C \to b$

Ans. $C \to abb$.

$S \to aAa$.

$A \to Sb \mid bcc$

T Q10: Remove useless symbols from the Grammar

$S \to aA \mid bB$

$A \to aA \mid a$

$B \to bB$

$D \to ab \mid Ea$

$E \to aC \mid d$

Ans: $A \to a$.

$D \to ab$

$E \to d$

$S \to aA$

$A \to aA$

Reduced $S \to aA$

$A \to aA$

$E \to d$

1 Q.: Eliminate useless symbol

$S \to aA \mid a \mid Bb \mid cC$

$A \to aB$

$B \to a \mid Aa$

$C \to cCD$

$D \to ddd$

Ans: $S \to aA \mid Bb \mid a$

$A \to aB$

$B \to Aa$

$B \to a$

Reduced

$S \to aA \mid Bb \mid a$

$A \to aB$

$B \to a \mid Aa$

Eliminating Unit productions

Q1: $S \to AB$ $\quad$ $D \to E$ $\quad$ $S \to \cancel{A}d$

$A \to a$ $\quad$ $E \to a$ $\quad$ $A \to \cancel{B}d$

$B \to C|b$ $\quad$ $B \to \cancel{C}d$

$C \to D$ $\quad$ $C \to d$

Ans: $S \to AB$

$A \to a$

$B \to a|b$.

Q2: $I \to a|b|Ia|Ib|I0|I1$

$F \to I \,|\, (E)$

$T \to F \,|\, T*F$

$E \to T \,|\, E+T$

Ans: $I \to a|b|Ia|Ib|I0|I1$

$F \to a|b|Ia|Ib|I0|I1|(E)$

$T \to a|b|Ia|Ib|I0|I1|(E)|T*F$

$E \to a|b|Ia|Ib|I0|I1|(E)|T*F|E+T$

G_3: $S \to A \mid bb$

$A \to B \mid b$

$B \to S \mid a$

$S \to B \mid b \mid bb$

$S \to a \mid b \mid bb$

$A \to$

Ans. $S \to b \mid bb \mid a$

$A \to a \mid b \mid bb$

$B \to bb \mid a \mid b$

Removal of ε-productions

Q1: $S \to aA$

$A \to b \mid \epsilon$

Ans: $S \to aA \mid a$

$A \to b$

Q2: $S \to ABAC$

$A \to aA \mid \epsilon$

$B \to bB \mid \epsilon$

$C \to c$

Ans. $A \to aA \mid a$

$B \to bB \mid b$

$S \to ABAC \mid BAC \mid ABC \mid AAC \mid AC \mid BC \mid C$

$C \to c$

Q3: $S \to aSa$

$S \to bSb \mid \epsilon$

Ans: $S \to aSa \mid aa$

$S \to bSb \mid bb.$

Q4: $S \to a \mid Xb \mid aYa$

$X \to Y \mid \epsilon$

$Y \to b \mid X$

Ans: $X \to Y \mid \epsilon$

$Y \to b \mid Y \mid \epsilon$

$S \to a \mid Xb \mid aYa \mid b \mid aa$

$X \to Y$

$Y \to b \mid X$

Q5. $\underbrace{(a+b)^*}_{A} bb \underbrace{(a+b)^*}_{A}$

A $S \to AbbA$

$A \to \cancel{aA} bA \mid \epsilon$

$A \to aA \mid bA \mid a \mid b$

$S \to AbbA \mid bb \mid Abb \mid bbA$

CHOMSKY NORMAL FORM (CNF)

Step1: Eliminate all unit, epsilon and useless symbols

Step2: Convert into Non-terminal → Exactly 2 non-terminal

Non-terminal → One terminal

Q1:

$S \rightarrow bA \mid aB$

$A \rightarrow bAA \mid aS \mid a$

$B \rightarrow aBB \mid bS \mid a$

Ans:

$S \rightarrow C_b A \mid AB$ $\quad C_b \rightarrow b$

$A \rightarrow C_b D \mid BS \mid a$ $\quad D \rightarrow AA$

$B \rightarrow AE \mid C_b S \mid a$ $\quad E \rightarrow BB$

Q2: Change the grammar into CNF

$S \rightarrow 1A \mid 0B$

$A \rightarrow 1AA \mid 0S \mid 0$

$B \rightarrow 0BB \mid 1$

Ans:

$S \rightarrow C_1 A \mid C_0 B$ $\quad C_1 \rightarrow 1 \quad C_0 \rightarrow 0$

$A \rightarrow C_1 D \mid C_0 S \mid 0$ $\quad D \rightarrow AA$

$B \rightarrow C_0 E \mid 1$ $\quad E \rightarrow BB$

52,000
12,500

Q3: Change into CNF

$S \to a | b | CSS$

Ans: $S \to a | b$

$S \to CD$

$D \to SS$

Q4: Convert into CNF

$S \to abSb | a | aAb$

$A \to bS | aAAb$

Ans: $S \to C_a C_b S C_b | a | C_a A C_b$

$A \to C_b S | C_a A A C_b$

$S \to AB$ $S \to a$

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

$B \to S C_b$ $D \to A C_b$

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

$A \to C_b S$ $A \to FG$

$F \to C_a A$ $G \to A C_b$

Q_5: Convert into CNF

$S \to bA \mid aB$

$A \to bAA \mid aS \mid a$

$B \to aBB \mid bS \mid b$

Ans:

$S \to C_b A \mid C_a B$

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

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

The pdts are

$S \to C_b A \mid C_a B$

$C_b \to b$

$C_a \to a$

$A \to C_b D$

$D \to AA$

$A \to C_a S \mid a$

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

$E \to BB$

GREIBACH NORMAL FORMS

Step1: Eliminate unit, null, ϵ pdns

step2: Make into CNF

Step3. $A \to a\alpha$ where $\alpha \in V_n^*$;

$A \to a$

Pdn are in the form

Lemma1: Let $G = (V, T, P, S)$ be a CFG. Let the set of A pdn be

$A \to A\alpha_1 | A\alpha_2 | \ldots A\alpha_r | \beta_1 | \beta_2 | \ldots | \beta_s$ where

(B_i's don't start with A)

1) Set of A_1 pdn are $A \to \beta_1 | \beta_2 | \ldots | \beta_s$

$A \to \beta_1 Z | \beta_2 Z | \ldots | \beta_s Z$

2) Set of Z pdn are $Z \to \alpha_1 | \alpha_2 | \ldots \alpha_r$

$Z \to \alpha_1 Z | \alpha_2 Z_2 | \ldots | \alpha_r Z$

Q_6: Convert into CNF

$S \to aAbB$, $A \to aA \mid a$, $B \to bB \mid b$

Ans: $A \to a$, $B \to b$ are in required form

$A \to C_a A$, $B \to C_b B$, $C_a \to a$, $C_b \to b$ are in required form

$S \to C_a A C_b B$

$S \to C_a C_1$

$C_1 \to A C_2$

$C_2 \to C_b B$

So the productions are

$S \to C_a C_1$	$A \to C_a A$	$C_b \to b$
$C_1 \to A C_2$	$B \to C_b B$	$A \to a$
$C_2 \to C_b B$	$C_a \to a$	$B \to b$

Q_7: Convert into CNF

$S \to nS \mid [S \supset S] \mid p \mid q$

Ans: $S \to p$, $S \to q$ are in required form

$S \to AS$ $A \to n$ are in required form

$S \to BSCDSB$ $S \to BC_1$, $C_1 \to SC_2$, $C_2 \to CC_3$

$C_3 \to DC_4$ $C_4 \to SB$

1) Write down the A pdn of CFG given below

$A \to aBD \mid bDB \mid c \qquad A \to AB \mid AD$

Ans: A pdn are those pdn that don't start with A

$A \to aBD \mid bDB \mid c$

$A \to aBDZ \mid bDBZ \mid cZ$

Z pdns are

$Z \to B \mid D$

$Z \to BZ \mid DZ$

Q2 Convert into GNF

$S \to AA \mid a \qquad A \to SS \mid b$

Step 1) Rename the variables S as A_1 and A as A_2. Grammar in CNF

$A_1 \to A_2 A_2 \mid a \qquad A_2 \to A_1 A_1 \mid b$

$A_1 \to A_2 A_2$, $A_1 \to a$, $A_2 \to b$ are in required form

$A_1 < A_2 \qquad 1 < 2$

$A_2 \rightarrow A_1 A_1 \mid b$ are not in required form since $A_2 \nless A_1$

$A_2 \rightarrow A_2 A_2 A_1 \mid a A_1 \mid b$

So A_2 pdn are

$A_2 \rightarrow a A_1 \mid b$

$A_2 \rightarrow a A_1 Z_2 \mid b Z_2$

Z_2 Pdn are

$Z_2 \rightarrow A_2 A_1 \mid A_2 A_1 Z_2$

The A_1 pdn are

$A_1 \rightarrow A_2 A_2 \mid a$

Substitute A_2 pdns in this

$A_1 \rightarrow a A_1 A_2 \mid b A_2 \mid a A_1 Z_2 A_2 \mid b Z_2 A_2 \mid a$.

The Z_2 pdns are

$Z_2 \rightarrow A_2 A_1 \mid A_2 A_1 Z_2$

Substitute A_2 in this

$Z_2 \rightarrow a A_1 A_1 \mid b A_1 \mid a A_1 Z_2 A_1 \mid b Z_2 A_1 \mid a A_1 A_1 Z_2 \mid b A_1 Z_2 \mid$

$a A_1 Z_2 A_1 Z_2 \mid b Z_2 A_1 Z_2$