PUMPING LEMMA FOR CFLs

FOR EVERY CFL L, THERE EXISTS $p \geq 0$ SUCH THAT EVERY $z \in L$ OF LENGTH $\geq p$ CAN BE DIVIDED INTO FIVE SUBSTRINGS

$z = uvwxy$ SUCH THAT

$vx \neq \epsilon$ AND $|vwx| \leq p$

AND FOR ALL $i \geq 0$, $uv^i wx^i y \in L$.

EVERY SUFFICIENTLY LONG STRING CAN BE DIVIDED INTO FIVE SEGMENTS SUCH THAT MIDDLE THREE ARE NOT TOO LONG. AND SECOND AND FORTH ARE NOT ~~NIL~~ BOTH NIL.

AND NO MATTER HOW ~~[illegible]~~ MUCH EXTRA IS ADDED BY PUMPING 2 AND 4TH SIMULTANEOUSLY THE STRING STAYS IN L.

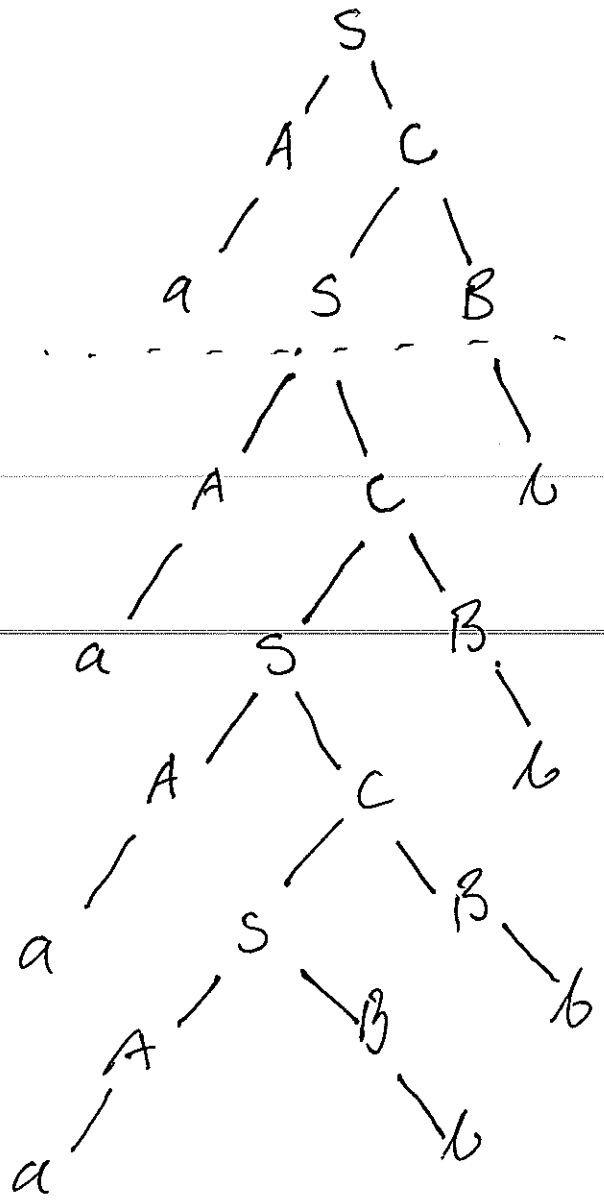
CONSIDER ~~PARSE~~ DERIVATION TREE FOR CHOMSKY GRAMMAR LANGUAGE

SHOWS ~~DERIVATIONS~~ PRODUCTIONS USED IN DERIVATION OF z FROM S.

EXAMPLE: $\{a^m b^m \mid m \geq 0\}$

$S \to AC \mid AB$, $A \to a$, $B \to b$, $C \to SB$

DERIVATION TREE FOR $z = a^4 b^4$



CONSIDER DERIVATION TREE FOR CHOMSKY GRAMMAR LANGUAGE

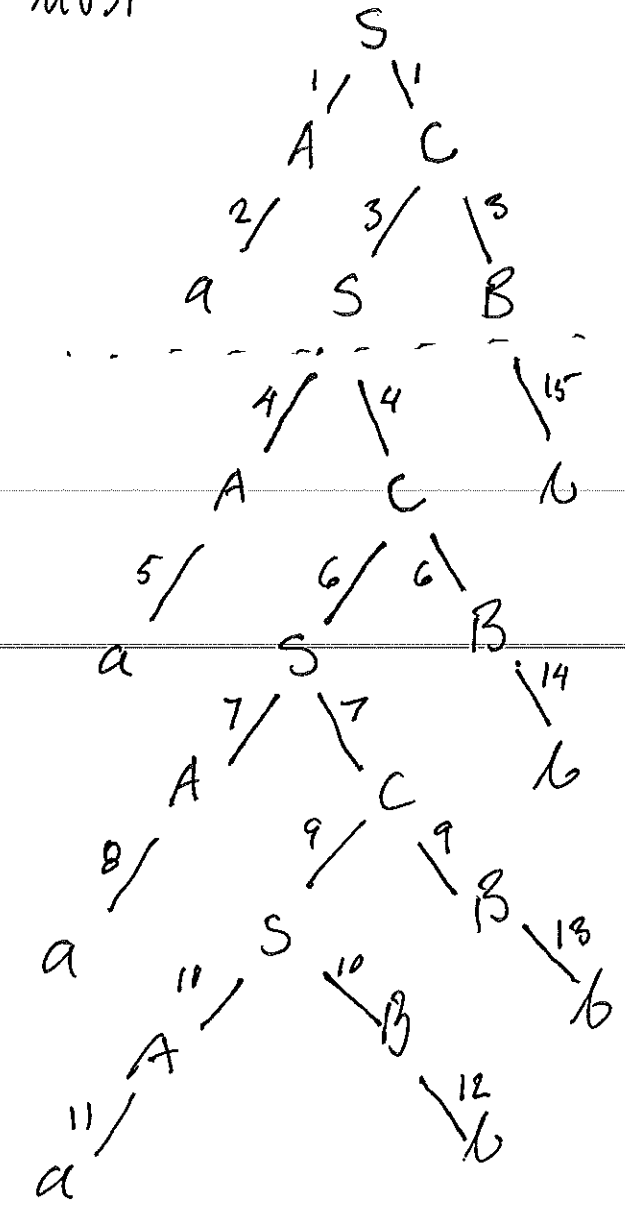
SHOWS PRODUCTIONS USED IN DERIVATION OF z FROM S.

EXAMPLE: $\{a^m b^m \mid m \geq 0\}$

$S \to AC \mid AB$, $A \to a$, $B \to b$, $C \to SB$

DERIVATION TREE FOR $z = a^4 b^4$

LEFT MOST



CONSIDER DERIVATION TREE FOR CHOMSKY GRAMMAR LANGUAGE

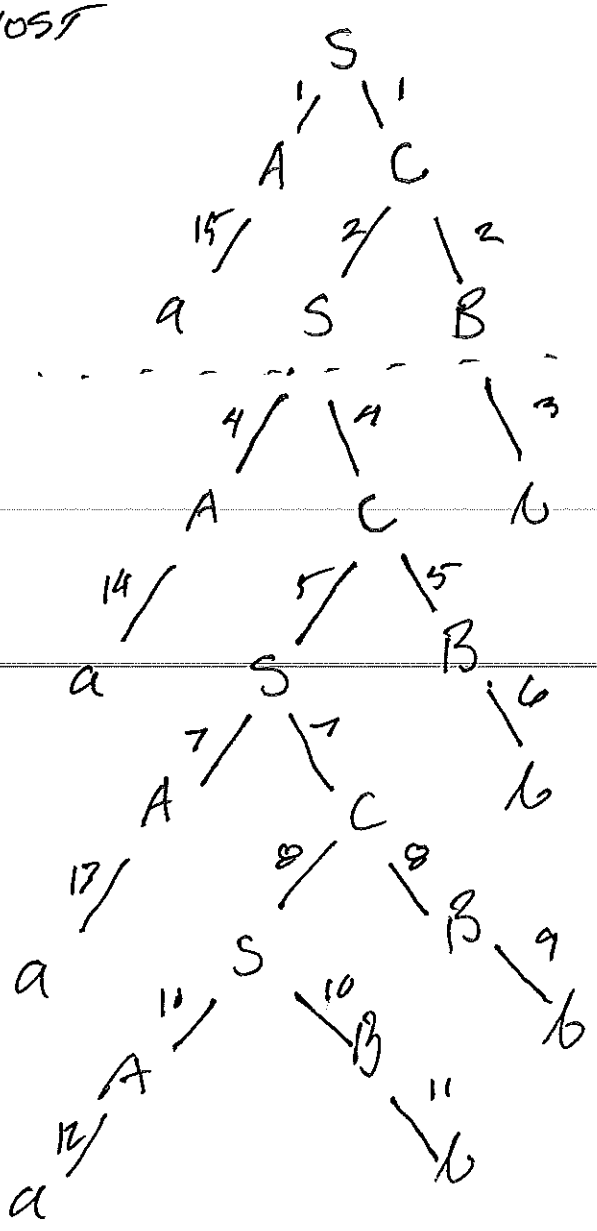
SHOWS PRODUCTIONS USED IN DERIVATION OF Z FROM S.

EXAMPLE: $\{a^n b^n \mid n \geq 0\}$

$S \to AC \mid AB$, $A \to a$, $B \to b$, $C \to SB$

DERIVATION TREE FOR $z = a^4 b^4$

RIGHT MOST



DERIVATION TREES FOR CHOMSKY GRAMMARS FOR LONG STRINGS MUST HAVE LONG PATHS

BECAUSE # OF SYMBOLS CAN AT MOST DOUBLE WHEN YOU GO DOWN ONE LEVEL.

RIGHT HAND SIDES OF PRODUCTIONS CAN'T HAVE MORE THAN 2 SYMBOLS.

OF SYMBOLS ON EACH LEVEL AT MOST 2X ABOVE LEVEL.

LEVEL	DEPTH	AT MOST THIS MANY SYMBOLS.
	0 TOP	1
	1	2
	2	4
	M	2^M

TO HAVE AT LEAST 2^M SYMBOLS AT BOTTOM MUST HAVE DEPTH M AT LEAST (M+1 LEVELS) I.E.

$\geq 2^M$ SYMBOLS $\Rightarrow$ DEPTH M

Length $\geq 2^{M+1}$ $\Rightarrow$ DEPTH $\geq M+1$

AT LEAST (M+1+1) LEVELS

PROOF.

LET G BE A GRAMMAR FOR L IN CHOMSKY NORMAL FORM.

LET $p = 2^{M+1}$

WHERE M = # OF VARIABLES IN G.

SUPPOSE $z \in L$ AND $|z| \geq p$

ANY DERIVATION TREE FOR z MUST HAVE A DEPTH AT LEAST M+1 PREVIOUS ARGUMENT.

CONSIDER LONGEST PATH IN TREE LENGTH IS AT LEAST M+1

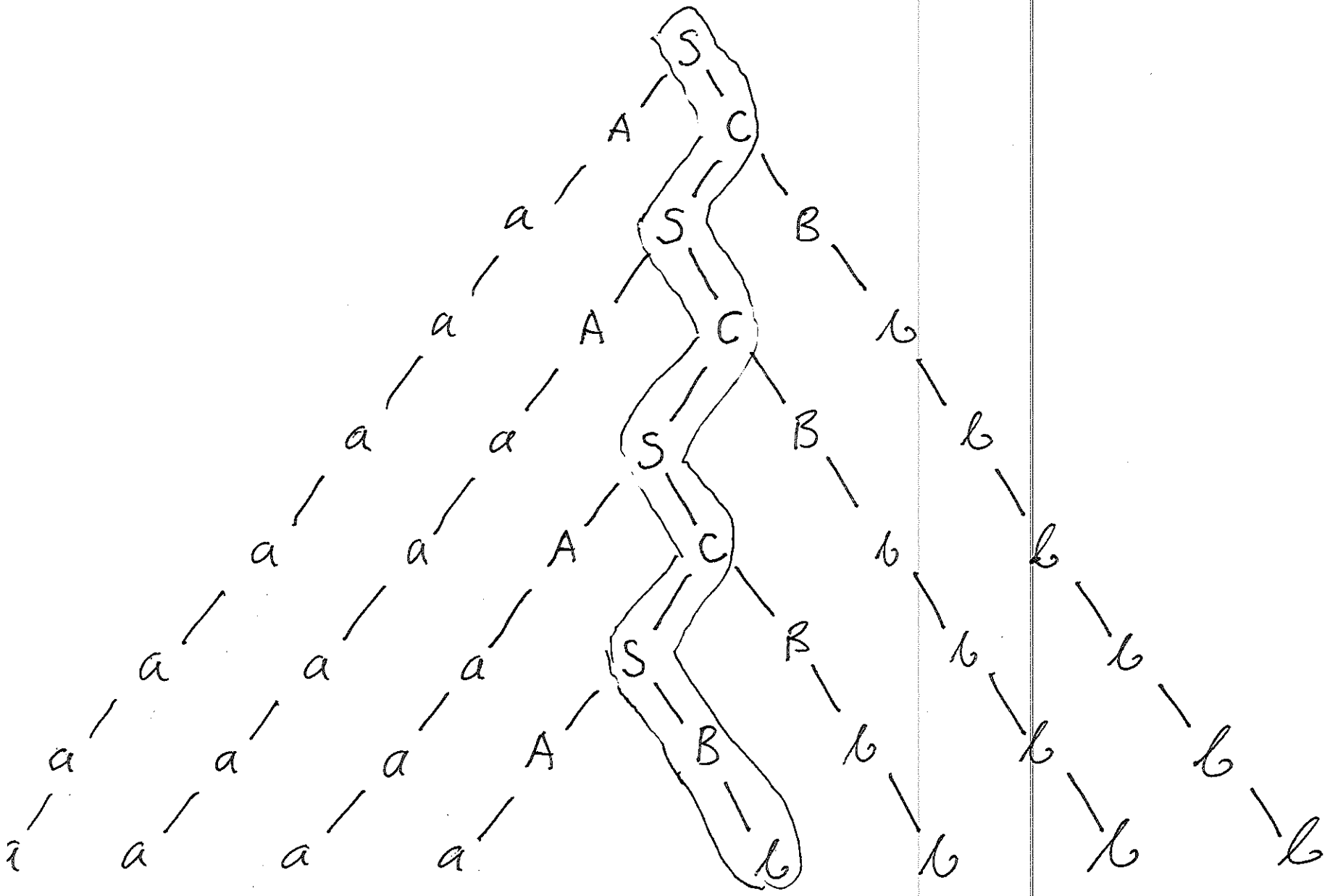
MUST CONTAIN M+1 OCCURRENCES OF VARIABLES. (EACH STEP ON PATH IS RULE)
BUT THERE ARE ONLY M

SO PIGEONHOLE PRINCIPLE SAYS SOME VARIABLE MUST OCCUR MORE THAN ONCE.

TAKE FIRST PAIR OF REPEATS FROM BOTTOM

(7)

ONE OF LONGEST PATHS.



n # distinct non terminals

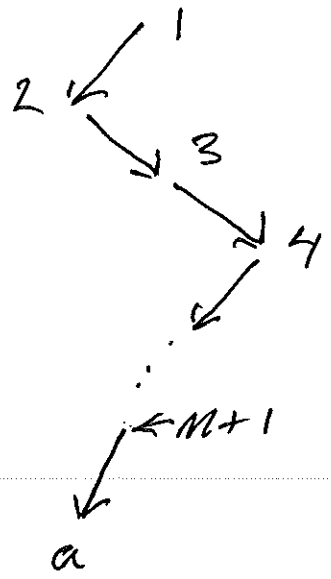
$|Z| \geq 2^{M+1}$ $\quad M$ = # VARIABLES

IMPLIES

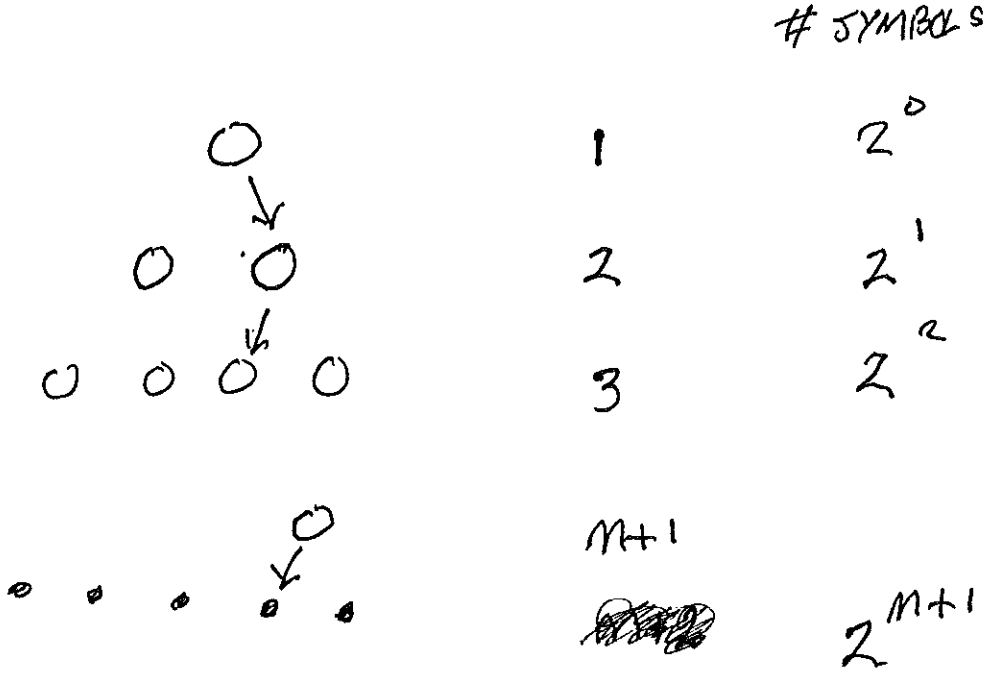
ANY DERIVATION TREE FOR Z HAS DEPTH $\geq M+1$

CONSIDER LONGEST PATH

IT MUST CONTAIN $M+1$ OCCURRENCES OF VARIABLES



PIGEON HOLE SOME VARIABLE MUST REPEAT.



○ Variable

• Terminal

SHORTEST POSSIBLE PATH TO DERIVE STRING WITH 2^{m+1} SYMBOLS

HAS $m+1$ VARIABLES ON IT

SO IF ONLY m VARIABLES POSSIBLE ONE MUST OCCUR AT LEAST TWICE.

SAY X IS VARIABLE WITH 2 OCCURRENCES

SPLIT z INTO SUBSTRINGS $uvwxy$

SUCH THAT w IS STRING GENERATED BY X_{LOW}

AND vwx IS STRING GENERATED BY X_{HIGH}

LET T BE SUBTREE ROOTED AT X_H AND

LET t " " " " X_L

BY REMOVING t FROM ORIGINAL TREE

AND REPLACING it WITH COPY OF T

WE GET VALID DERIVATION TREE FOR

$$uv^2wx^2y$$

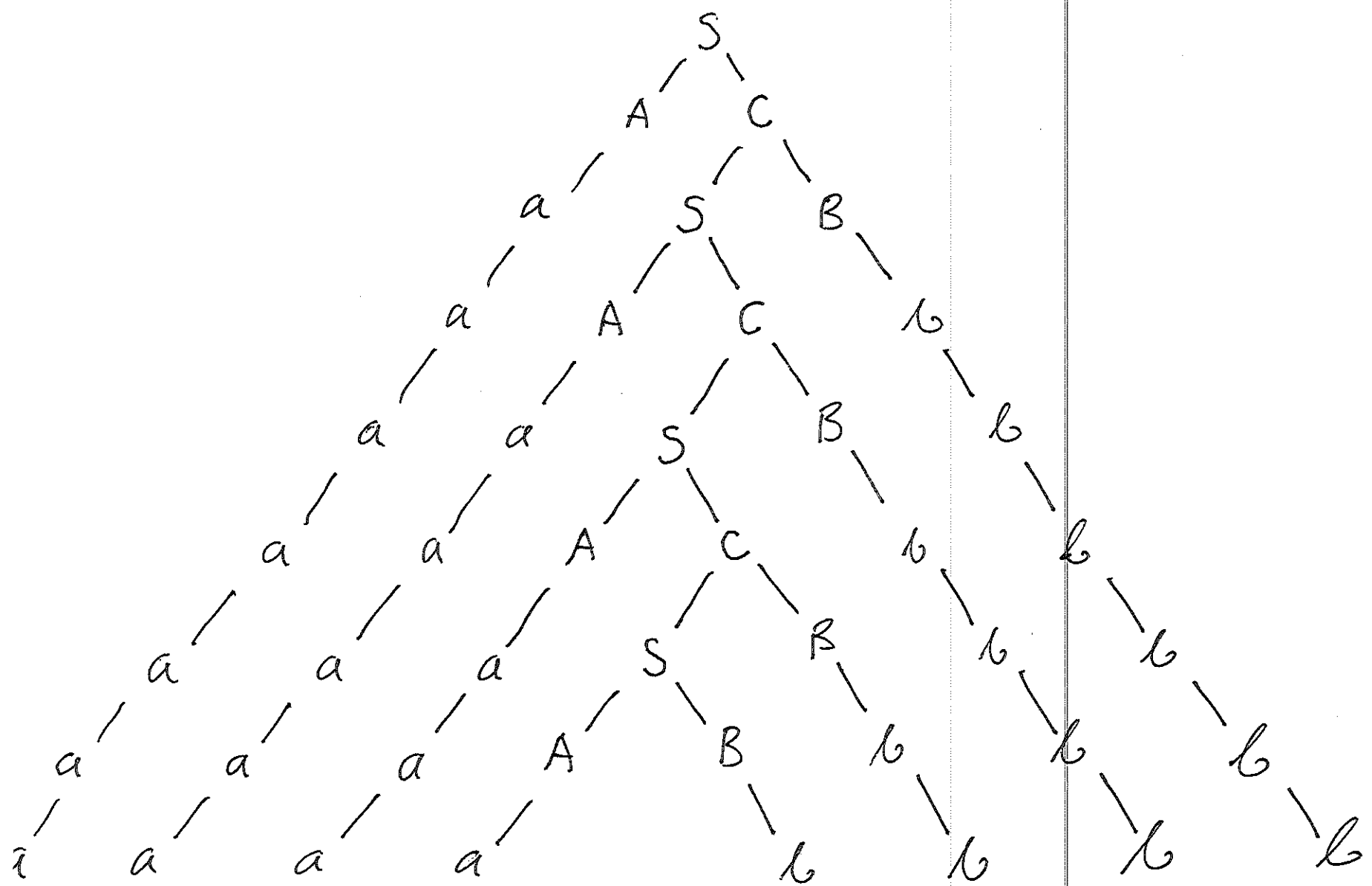
CAN DO THIS ARBITRARY # OF TIMES

$$uv^{i'}wx^{i'}y \qquad i'>1$$

FOR $i=0$ REPLACE T WITH t.

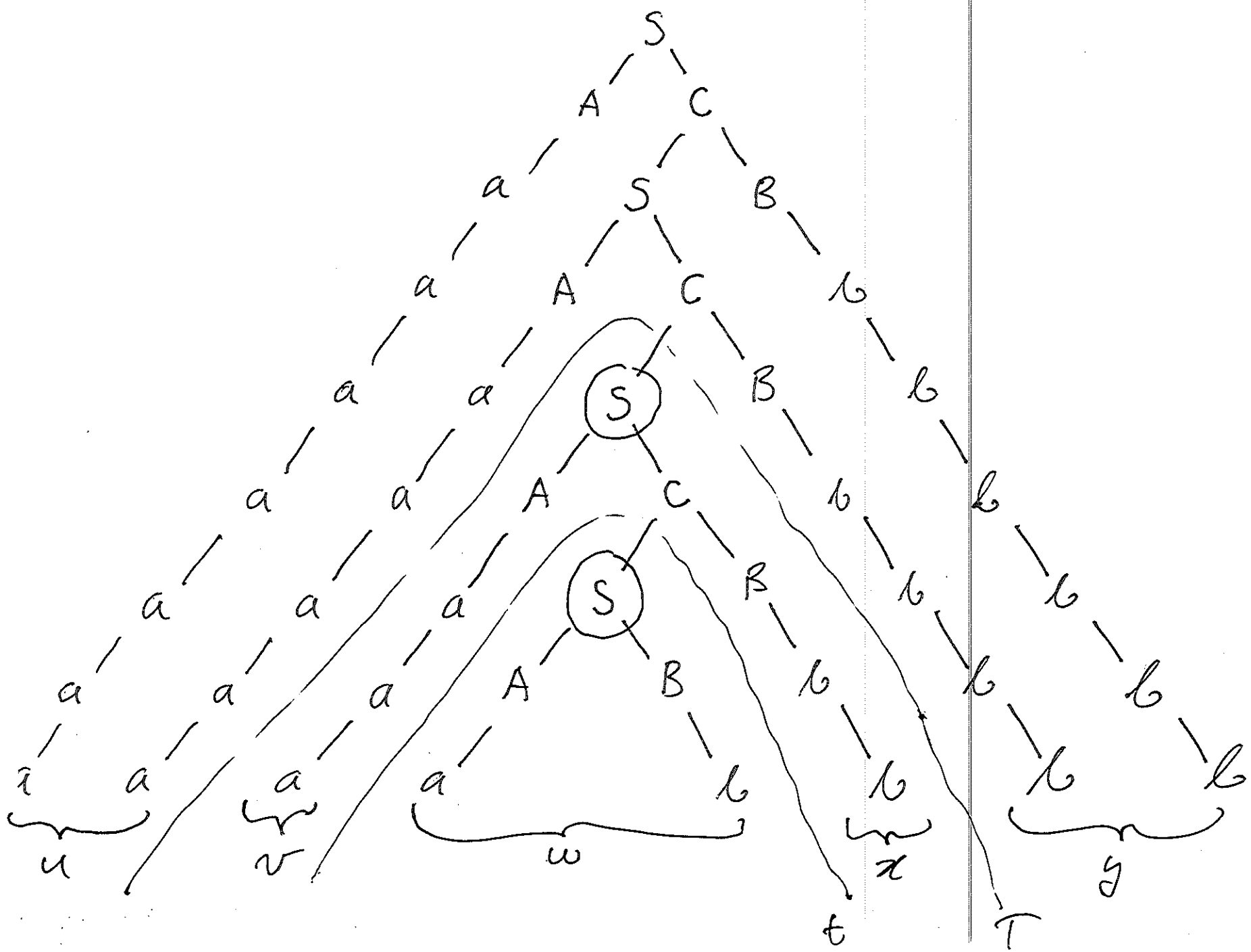
Q.E.D.

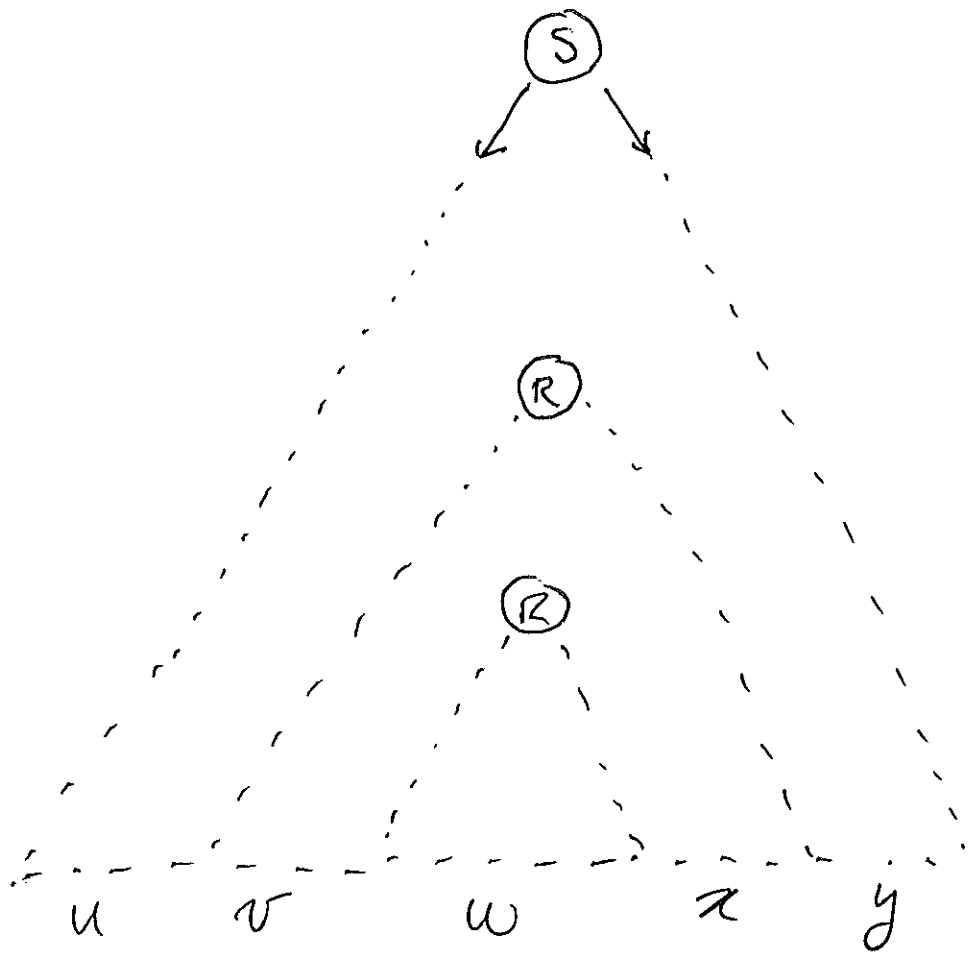
KEEP TRACK OF SYMBOLS GENERATED BY PUPPETEERING TERMINALS

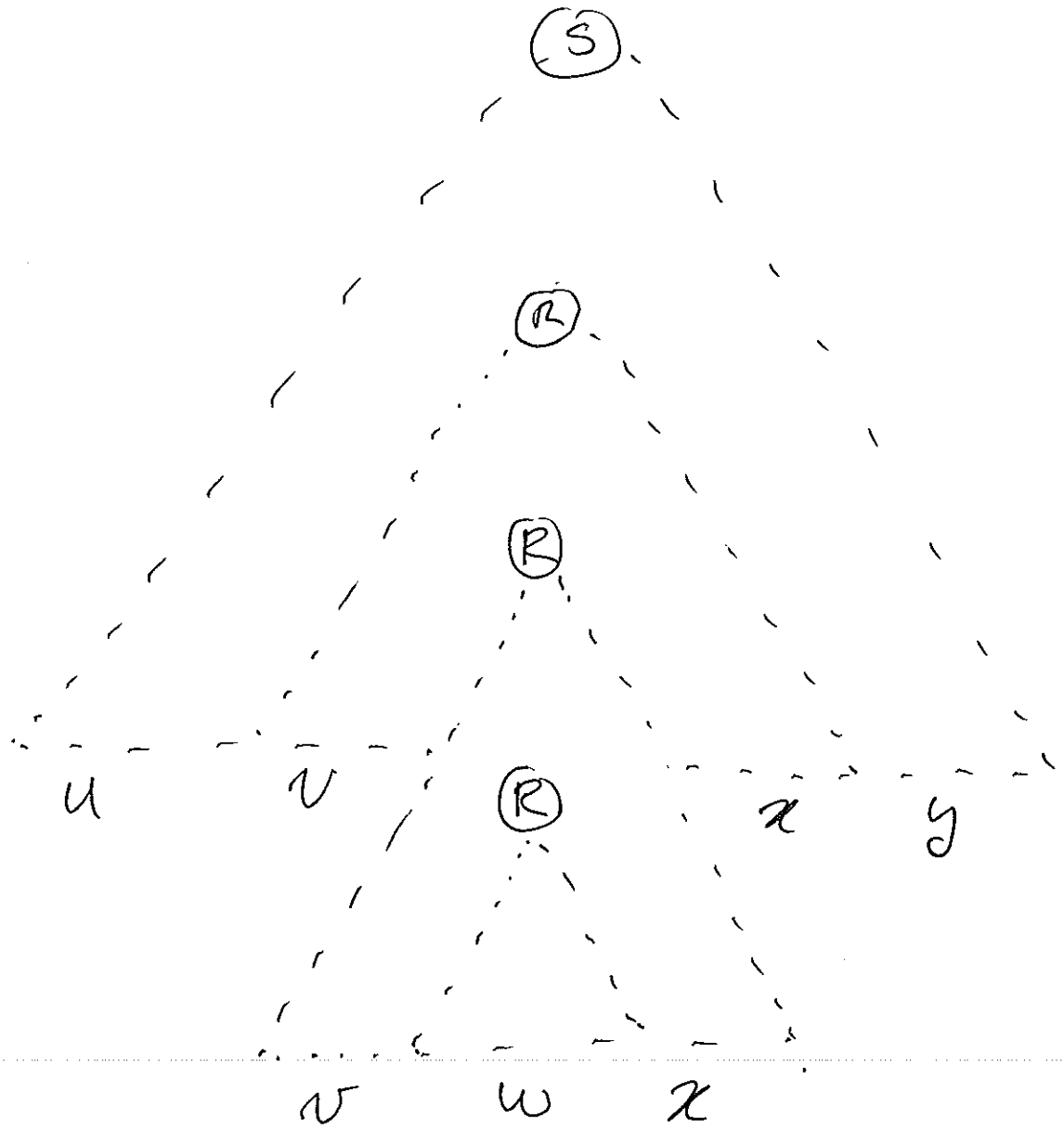


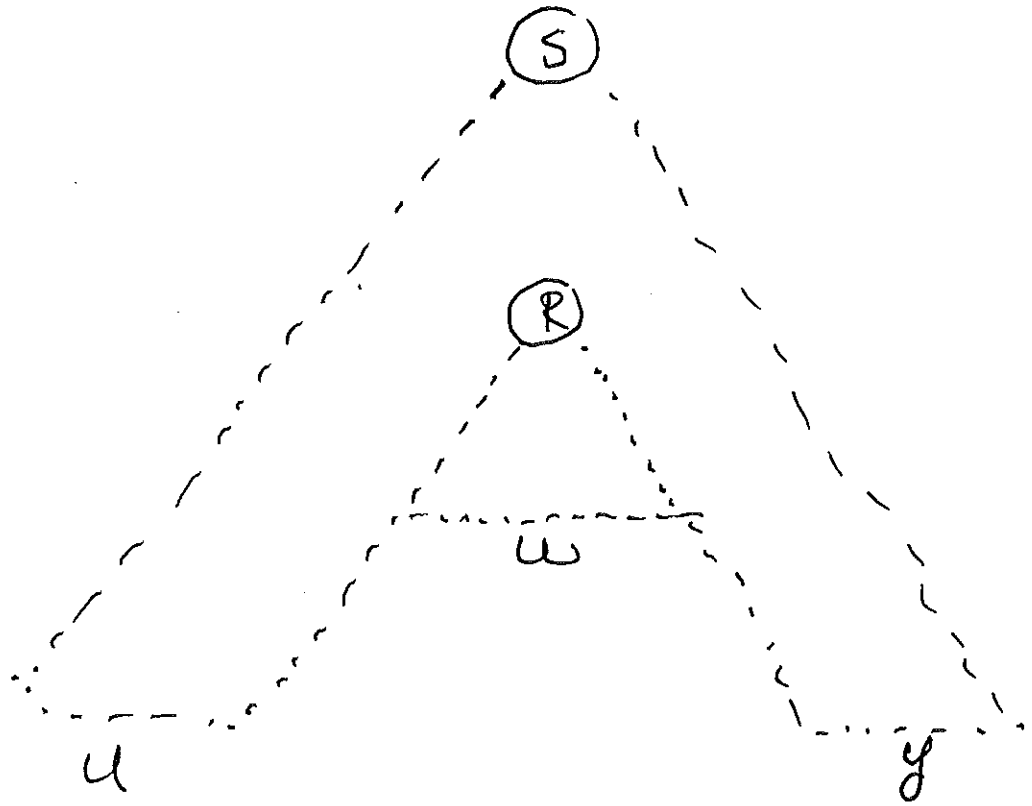
FIRST PAIR OF REPEATS FROM THE BOTTOM

KOZEN
R. 151




S
R
R
u
v
w
x
y


S
R
R
R
u
v
x
y
v
w
x


S
R
w
u
y

TO PROVE L NOT CFL. SHOW.

FOR ALL $p \geq 0$ THERE EXISTS $s \in L$
OF LENGTH AT LEAST p
SUCH THAT
FOR ALL $uvwxy$ WITH $z = uvwxy$
$vx \neq \varepsilon$
$|vwx| \leq p$

THERE EXISTS $i \geq 0$
SUCH THAT
$uv^i wx^i y \notin L$

---

1. DRAGON PICKS $p \geq 0$
2. YOU PICK $s \in L$ WITH $|s| \geq p$
3. DRAGON PICKS STRINGS u, v, w, x, y SUCH THAT $s = uvwxy$, $|vx| > 0$ AND $|vwx| \leq p$.

---

4. YOU PICK $i \geq 0$.
IF $uv^i wx^i y \notin L$ YOU WIN!

TO PROVE L NOT CFL SHOW YOU CAN WIN THE GAME REGARDLESS OF WHAT u, v, w, x, y DRAGON PICKS IN STEP 3.

$L = \{ a^n b^n c^n \mid n \geq 0 \}$ EX 1

$L = \{ a^n b^n a^n \mid n \geq 0 \}$ EX 2

$L : \{ w c w \mid w \in \{a, b\}^* \}$ EX 3

SIPSER

KUZEN

EX 1

$L = \{ a^n b^n c^n \mid n \geq 0 \}$

DEAMON PICKS p

WE PICK $s = a^p b^p c^p$

NOTE $s \in L$ AND $|s| = 3p \geq p$

DEAMON PICKS u, v, w, x, y SUCH THAT

$uvwxy = a^p b^p c^p$

$|vwx| \leq p$

$vx \neq \varepsilon$

WE PICK $i = 2$

THEN $uv^2wx^2y \notin L$ SO L IS NOT CFL

PROOF:

CASE I. $|vwx|$ IS ONLY a's OR ONLY b's OR ONLY c's
THEN THERE WILL BE MORE OF ONE OF THE LETTERS THAN THE OTHER TWO.

CASE II $|vwx|$ STRADDLES a's & b's OR b's & c's
THEN EITHER a's OR b's OR BOTH ARE INCREASED BUT NOT c's; OR b's OR c's OR BOTH ARE INCREASED BUT NOT a's.

OR

ANY OTHER ARGUMENT THAT IS VALID.

EX 2

$L = \{ a^n b^n c^n \mid n \geq 0 \}$

p

$s = a^p b^p c^p \quad s \in L \quad \text{AND} \quad |s| \geq p$

~~$a^p b^p c^p$~~

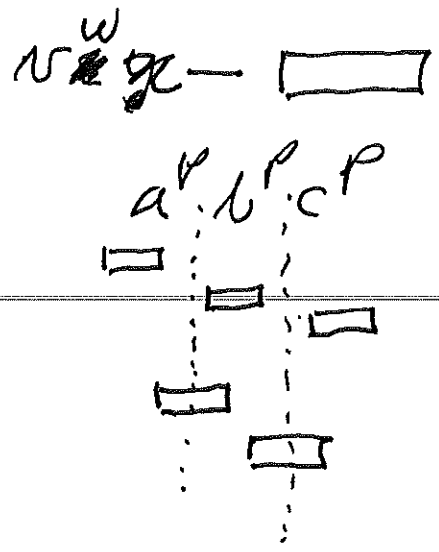
$uvwxy = s$

$vwx \leq p$

$vx \neq \varepsilon$

$i = 2$

~~$uvwxy$~~ $\quad uv^2wx^2y \notin L$



} more of one than the others a_s b_s c_s

more a_s or b_s than c_s

more b_s or c_s than a_s

$A = \{ a^n b^n a^n \mid n \geq 0 \}$ EKL

DRAGON PICKS k

PICK $z = a^k b^k a^k$

NOTE $z \in A$ AND $|z| = 3k \geq k$.

DRAGON PICKS u, v, w, x, y

SUCH THAT $z = uvwxy$

$|vx| \geq 1$ $vx \neq \epsilon$

$|vwx| \leq k$

PICK $z' = z$

CASES

EITHER v OR x HAS (ONE OF THEM HAS MIXTURE) ^ AT LEAST ONE a AND AT LEAST ONE b

THEN uv^2wx^2y IS NOT OF FORM $a^*b^*a^*$

$ab \ldots ab$

$ba \ldots ba$

SO $uv^2wx^2y \notin A$.

---

v AND x ONLY CONTAIN a's

THEN uv^2wx^2y HAS a's INCREASED BUT NOT b's

SO NOW HAS MORE ~~THAN TWICE AS MANY~~ a's THAN b's.

SO ..... $\notin A$

---

v AND x ONLY CONTAIN b's

SIMILAR TO ABOVE

NOW FEWER ~~THAN TWICE AS MANY~~ a's AS b's.

SO ..... $\notin A$

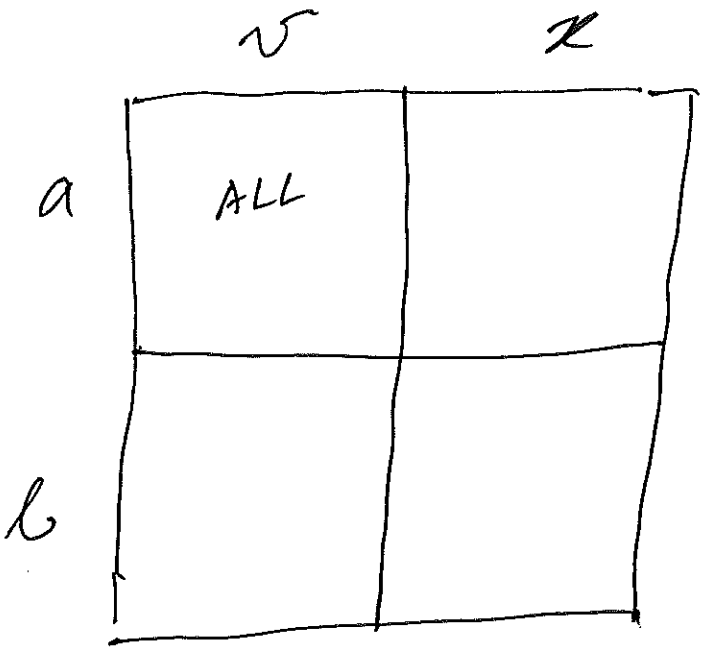
ONE OF v OR x CONTAINS ONLY a's AND THE OTHER ONLY b's

$a^m b^m a^n$ OR $a^m b^n a^n$

AND uv^2wx^2y CAN NOT BE OF FORM $a^m b^m a^m$

SO ..... $\notin A$. Q.E.D. MORE LEFT a's THAN RIGHT a's OR VISA VERSA.

EX 2



v	a·b		a	b	a	b
x		a·b	a	b	b	a

EX 3

$D = \{ ww \mid w \in \{a,b\}^* \}$

NOTE $CFL \cap REG \to CFL$

CONSIDER

$D' = D \cap L(a^*b^*a^*b^*)$ SO IF D IS CFL THEN D' MUST BE

$= \{a^m b^m a^m b^m \mid a, b \geq 0\}$ BUT D' IS NOT

p

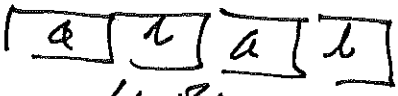
$s = a^p b^p a^p b^p$

$s = uvwxy$

$|vwx| \leq p$

$vx \neq \varepsilon$

$z' = z$



4 BLOCKS

I. ONE OF v OR x CONTAINS BOTH a's AND b's. THEN PUMPING DESTROYS FORM $a^*b^*a^*b^*$

II v AND x BOTH FROM SAME BLOCK THEN MORE a's ON ONE SIDE THAN THE OTHER OR MORE b's ON ONE SIDE THAN THE OTHER

III v CONTAINS ONE SYMBOL x THE OTHER SYMBOL. SO BLOCKS MUST BE ADJACENT. WILL CREATE IMBALANCE FOR THE # a's AND/OR b's FOR THE TWO HALVES.

EX3