

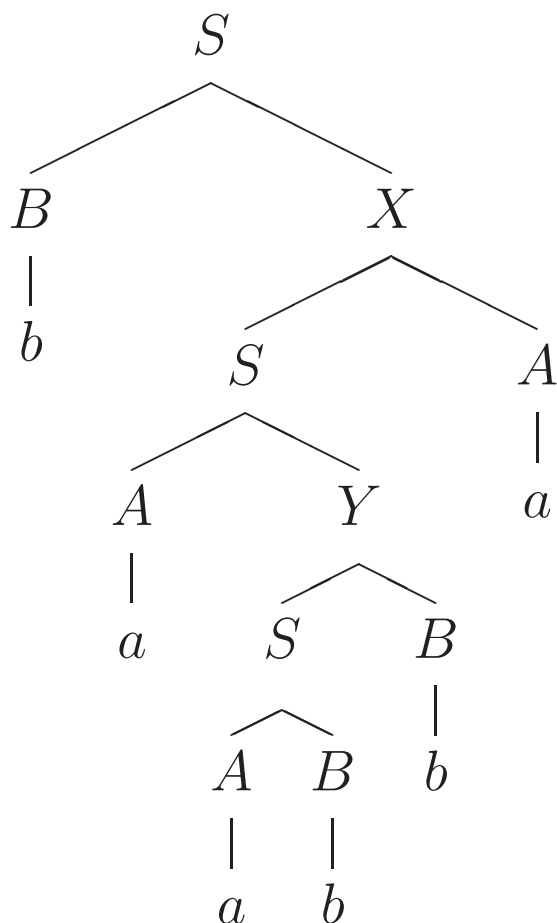
復習：導出木

注意：来週の間中間試験は持ち込み不可です

a と b を同じ数含む長さ 1 以上の記号列を生成する文法のチョムスキー標準形の生成規則

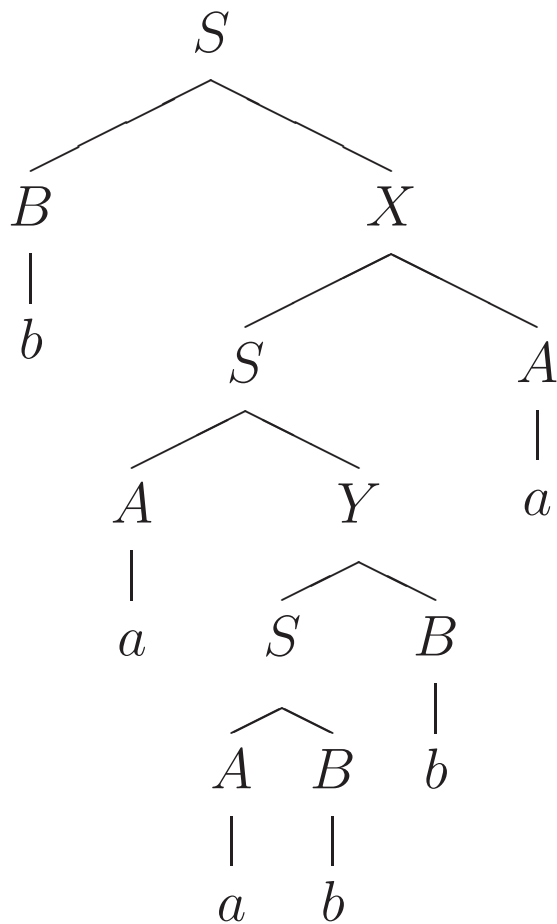
$$P = \{S \rightarrow AB|BA|AY|BX|SS, \\ X \rightarrow SA, Y \rightarrow SB, A \rightarrow a, B \rightarrow b\}$$

$baabba$ の導出木



上記のような表記法を導出木と呼ぶ。

文脈自由文法のポンプの補題の例



上の例で $u = b$, $v = a$, $w = ab$, $x = b$, $y = a$ と置くと、まんなかにある S を書き換える操作を繰り返して、

$$uv^iwx^iy = ba^iabb^ia$$

も導出できることがわかる。このことを一般的に書くと...

文脈自由文法のポンプの補題

L が文脈自由言語ならば、ある n が存在し、 $|z| \geq n$ である L の記号列 z について以下の条件を満たす記号列 u, v, w, x, y が存在する

1. $z = uvwxy$
2. $|vx| \geq 1$
3. $|vwx| \leq n$
4. すべての $i \geq 0$ について $uv^iwx^iy \in L$

証明：

$G: L \setminus \{\epsilon\}$ を生成するチョムスキー標準形

m : G の非終端記号の数

$$n = 2^m$$

導出木の高さを、開始記号から終端記号に至る枝の数の最大値とする。高さが h の導出木を持つ記号列の長さは 2^h 以下である。従って $|z| \geq n$ なら z の導出木の高さは m 以上である。導出木の開始記号から終端記号に至る経路で最も長いものには $m + 1$ 個以上の非終端記号がある。2 回以上現れる非終端記号のうち終端記号に最も近いものを A として、

$$S \xRightarrow{*} uAy,$$

$$A \xRightarrow{*} vAx,$$

$$A \xRightarrow{*} w,$$

と u, v, w, x, y を定義すると (教科書 85 ページの図 5.3 を参照)、

1. $z = uvwxy$
2. 導出木の上にある A は 2 つの子を持つので $|vx| \geq 1$
3. (終端記号近い非終端記号を選んだので) 導出木の上にある A から終端記号まで枝の数は m 以下である。従って $|vwx| \leq 2^m = n$
4. $A \xRightarrow{*} vAx$ の導出を i 回繰り返すと uv^iwx^iy を導出できる

ポンプの補題の使用例

$L = \{a^m b^m c^m \mid m \geq 1\}$ は文脈自由文法ではない。

ある n が存在して

1. $a^n b^n c^n = uvwxy$,
2. $|vx| \geq 1$,
3. $|vwx| \leq n$,
4. $uv^i wx^i y \in L$

とできたとして矛盾を導く。条件 3 より vx の中は a と b だけ含むか b と c だけ含む。従って uv^2wx^2y は a と b と c の数が等しくないので L に含まれず矛盾。

演習問題 50 $\{a^m b^m c^i \mid m \geq 1, i \leq m\}$ が文脈自由言語ではないことを証明せよ。

演習問題 51 $\{ww \mid w \in \{a, b\}^+\}$ が文脈自由言語ではないことを証明せよ。

問題 52 今日の授業でわかりにくい所や要望を書いて下さい

演習問題解説

問題 47 文脈自由文法 $G = (\{S, A, B\}, \{a, b\}, P, S)$,
 $P = \{S \rightarrow \epsilon | ASB, A \rightarrow a | aAS, B \rightarrow bb | A | SbS\}$ に対して、 $L(G)$ から ϵ を除いた言語を生成する文脈自由文法の CNF を (N, T, P, S) の形で書け

まず $L(G)$ から ϵ を除いた言語を生成する文法を前回やった方法に従って作る。

1. P から $S \rightarrow \epsilon$ を除いた生成規則の集合を P_1 とする
2. ϵ 生成記号は S だけ
3. 新しい生成規則の集合は

$$P_2 = \{S \rightarrow ASB | AB, A \rightarrow a | aAS | aA, \\ B \rightarrow bb | A | SbS | bS | Sb | b\}$$

P_2 を CNF に変換する

1. $U(S) = U(A) = \emptyset, U(B) = A$. 従って $B \rightarrow A$ を除き代わりに $B \rightarrow a|aAS|aA$ を追加して

$$P_3 = \{S \rightarrow ASB|AB, A \rightarrow a|aAS|aA, \\ B \rightarrow bb|SbS|bS|Sb|b|a|aAS|aA\}$$

2. $X_a \rightarrow a, X_b \rightarrow b$ を付け加えて

$$P_4 = \{S \rightarrow ASB|AB, A \rightarrow a|X_aAS|X_aA, \\ B \rightarrow X_bX_b|SX_bS|X_bS|SX_b|b|a|X_aAS|X_aA, \\ X_a \rightarrow a, X_b \rightarrow b\}$$

3. $Y_1 \rightarrow SB, Y_2 \rightarrow AS, Y_3 \rightarrow X_bS$ を付け加えて

$$P_5 = \{S \rightarrow AY_1|AB, A \rightarrow a|X_aY_2|X_aA, \\ B \rightarrow X_bX_b|SY_3|X_bS|SX_b|b|a|X_aY_2|X_aA, \\ X_a \rightarrow a, X_b \rightarrow b, \\ Y_1 \rightarrow SB, Y_2 \rightarrow AS, Y_3 \rightarrow X_bS\}$$

求める CNF は $(\{S, A, B, X_a, X_b, Y_1, Y_2, Y_3\}, \{a, b\}, P_5, S)$

問題 48 アルファベット $\{a, b\}$ 上の言語 $L_{48} = \{ww \mid w \in \{a, b\}^*\}$ が正規言語ではないことを証明せよ。

L_{48} を受理する状態数 n の DFA が存在したとして矛盾を導く。 $a^n b^n a^n b^n = wxy$, $|x| \geq 1$, $|wx| \leq n$ と分けると x は最初の a^n の中にある。従って $wy = a^{n-|x|} b^n a^n b^n$ だが、 $wy \notin L_{48}$ で矛盾