計算理論　第13回
第7章：
文脈自由言語の性質 (2/3)

中川　博之

h-nakagawa@okayama-u.ac.jp

本日の概要

- 第7章：文脈自由言語の性質
  - テキスト：p.294～
  - 7.1.5 チョムスキーの標準形
  - 7.2 文脈自由言語の反復補題

- 重要概念
  - Chomsky標準形，（文脈自由言語の）反復補題

7.1.5 チョムスキーの標準形

チョムスキー標準形（Chomsky Normal Form：CNF）

- 生成規則を以下の形に限定したCFG
  - A→BC　（A,B,Cは変数）
  - A→a　　（Aは変数，aは終端記号）
  - また，無用な記号を含まない

- Chomsky
  - CFGを提案した言語学者

定理7.16

- 文法G：ε以外の列を少なくとも1つ生成するCFGとすると,
  $L(G_1) = L(G) - \{\varepsilon\}$を満たす
  チョムスキー標準形文法G_1が存在
- 証明略（教科書参照）
- どのようなCFGでもチョムスキー標準形で文法を記述可能なことを意味する
  - $L(G)=\{\varepsilon\}$の場合を除く

チョムスキー標準形への変換：変換前の文法

変換前のCFG G

- 仮定：ε-規則，単位規則，無用な記号はない
  - ここまでで示した方法で除去しておく

Gの生成規則は以下のいずれかの形となる

- 生成規則の形1： $A \rightarrow a$ （$a \in T$）
  - 右辺は1つの終端記号だけ
- 生成規則の形2： $A \rightarrow B_1 B_2 B_3 \ldots B_k$
  - $B_i \in V \cup T$, $k \geq 2$

チョムスキー標準形への変換: 前処理

- 生成規則の右辺が長さ2以上のもの
→ 右辺に<u>終端記号を含まないように変換</u>する

- 例: 生成規則 X→aXbYに対して
  - X→AXBY
  - A→a
  - B→b

- 一般的な変換方法
  - 右辺中の各終端記号aを対応する変数Aで置き換える
  - 生成規則A→aを追加

チョムスキー標準形への変換（1/2）

- 右辺の長さ=1のもの：$A \rightarrow a$
  - そのまま生成規則として採用
- 右辺の長さ=2のもの：$A \rightarrow B_1 B_2$
  - そのまま生成規則として採用
- 右辺の長さが3以上のもの：$A \rightarrow B_1 B_2 \ldots B_k$（$k \geq 3$）
  - 右辺の長さが2になるように変換(次スライド)

チョムスキー標準形への変換(2/2)

- 右辺の長さが3以上のもの: $A \rightarrow B_1 B_2 \ldots B_k$ ($k \geq 3$)
  - $C_1, C_2, \ldots, C_{k-2}$を導入して以下のように変換
  - $A \rightarrow B_1 C_1$
  - $C_1 \rightarrow B_2 C_2, \ldots$
  - $C_{k-2} \rightarrow B_{k-1} B_k$

- 例: $A \rightarrow B_1 B_2 B_3$
  - $A \rightarrow B_1 C_1$
  - $C_1 \rightarrow B_2 B_3$
  - C_1を新たに導入

例7.15

- 例7.12の文法をCNFに
  - E → E+T|T*F|(E)|a|b|Ia|Ib|I0|I1
  - T → T*F|(E)|a|b|Ia|Ib|I0|I1
  - F → (E)|a|b|Ia|Ib|I0|I1
  - I → a|b|Ia|Ib|I0|I1
- まず，前処理として，終端記号を変数に書き換え，終端記号を導出する規則を追加
  - A→a, B→b, Z→0, O→1, P→+, M→*, L→(, R→)
    - 長さ1の本体にしか出現しない終端記号については不要だが，今回はすべての終端記号が長さ2以上の本体に現れるために必要

例7.15

- 前処理後
  - E → EPT|TMF|LER|a|b|IA|IB|IZ|IO
  - T → TMF|LER|a|b|IA|IB|IZ|IO
  - F → LER|a|b|IA|IB|IZ|IO
  - I → a|b|IA|IB|IZ|IO
  - A→a, B→b, Z→0, O→1, P→+, M→*, L→(, R→)
- 本体の長さ3以上の規則を分割するためにC_iを適宜追加
  - E→EPTをE→EC_1, C_1→PTなど

例7.15

- 得られたCNF
  - $E \to EC_1 \mid TC_2 \mid LC_3 \mid a \mid b \mid IA \mid IB \mid IZ \mid IO$
  - $T \to TC_2 \mid LC_3 \mid a \mid b \mid IA \mid IB \mid IZ \mid IO$
  - $F \to LC_3 \mid a \mid b \mid IA \mid IB \mid IZ \mid IO$
  - $I \to a \mid b \mid IA \mid IB \mid IZ \mid IO$
  - $A \to a$, $B \to b$, $Z \to 0$, $O \to 1$, $P \to +$, $M \to *$, $L \to ($, $R \to)$
  - $C_1 \to PT$
  - $C_2 \to MF$
  - $C_3 \to ER$

7.2 文脈自由言語の反復補題

文脈自由言語の反復補題

- どのCFL Lでも必ず有する性質P(L)
  - 必要条件
  - （十分条件ではない）

- LがCFLならばP(L)は真

  対偶を取ると

  P(L)が偽ならば，LはCFLではない

- 言語がCFLではないことを示すための手段

7.2.1 構文木の大きさ

復習：チョムスキー標準形 (CNF)

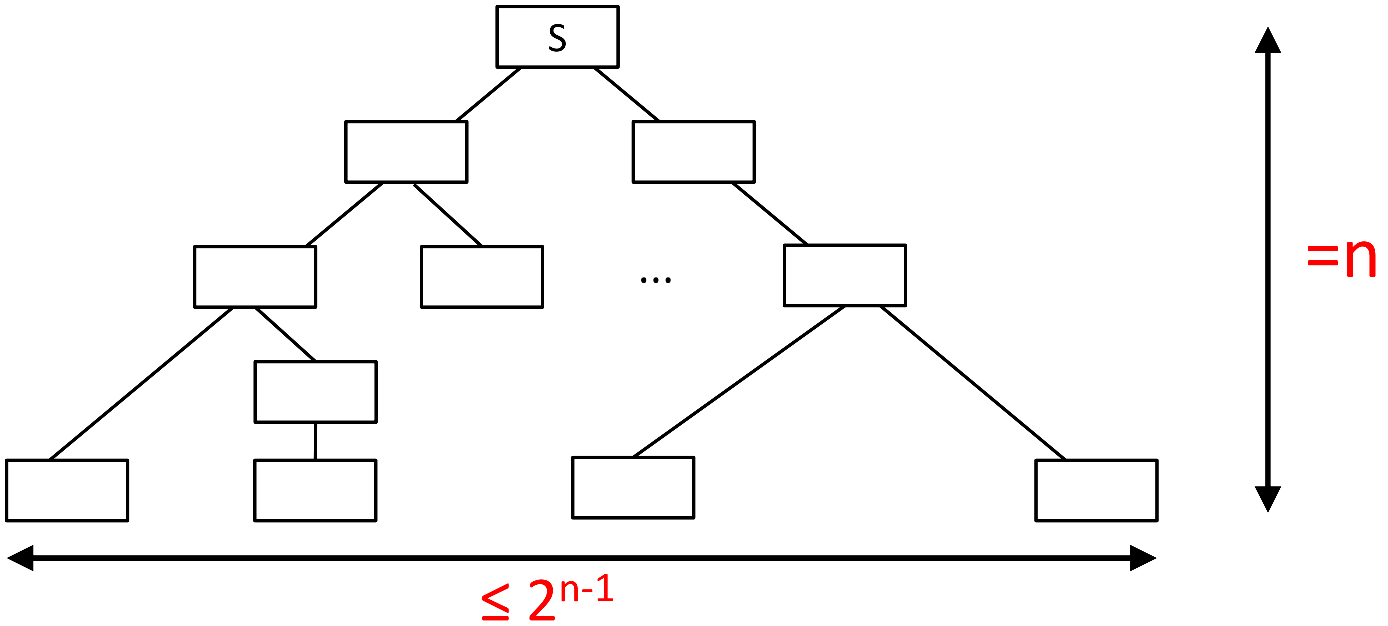
生成規則を以下の形に限定したCFG

- A→BC　　（A, B, Cは変数）
- A→a　　（Aは変数，aは終端記号）
- 構文木が2分木になるのが特徴

定理7.17：チョムスキー標準形の構文木の大きさ

- G=(V, T, P, S)：チョムスキー標準形の文法
- w∈L(G)
- n：wの構文木の最長路の長さ

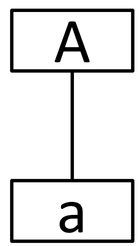
とすると，　$|w| \leq 2^{n-1}$



定理7.17：証明

- 基礎：n=1
  - 根の子が葉のみの場合に相当
  - チョムスキー標準形なので，生成規則はA→a
  - 葉の数は1つだけなので，|w| = 1

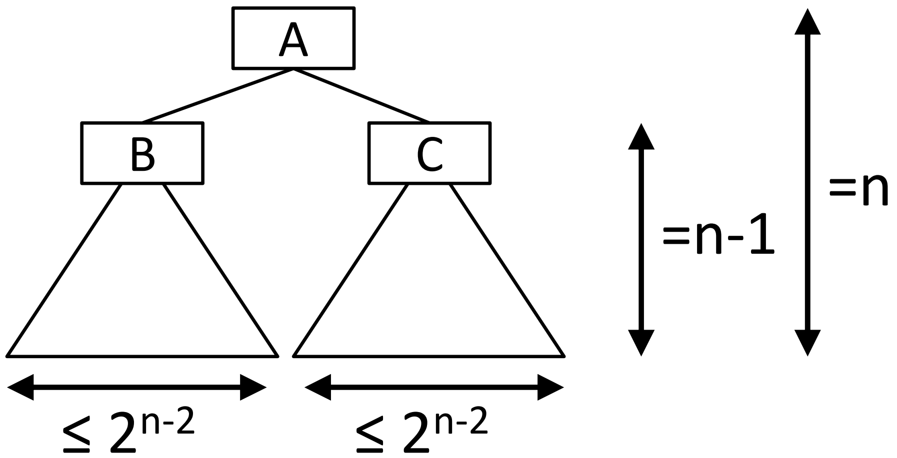
→ $|w| = 1 \leq 2^{n-1} = 2^0 = 1$ 不等式が成立



定理7.17:証明

- 帰納: n>1
  - 根に該当する生成規則: A→BC
  - Bを根とする部分木の最長路の長さ:高々n-1
    - 帰納法の仮定より, 葉の数は高々2^{n-2}
  - Cを根とする部分木も同様

→ $|w| \leq 2^{n-2} + 2^{n-2} = 2^{n-1}$　不等式が成立



7.2.2 反復補題の表現

テキスト p.303

定理7.18：CFLの反復補題

- L：任意のCFL
- n：Lによって決まる定数
- z：Lに属する列で$|z| \geq n$であるもの

以下の条件を満たすようなzの分解$z=uvwxy$が存在

1. $|vwx| \leq n$
2. $vx \neq \varepsilon$ (vとxが同時にεになることはない)
3. 任意の$i \geq 0$に対し，$uv^iwx^iy \in L$ が成立

証明：準備

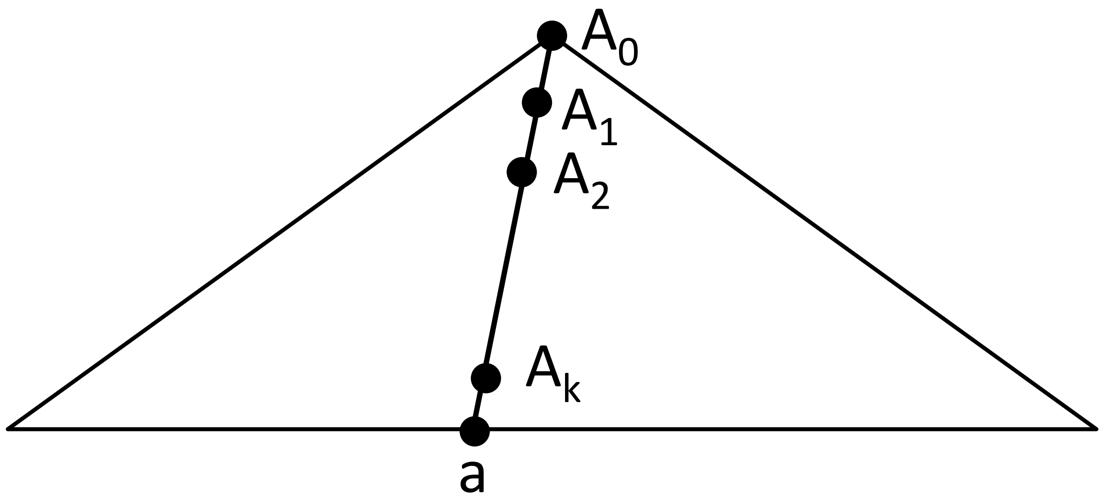
- G：Lを生成するチョムスキー標準形の文法とする
- m：Gの変数の数
- $n = 2^m$となるnを選ぶ
- z：Lに属する列で $|z| \geq n$ であるもの

→ 定理7.17（$|w| \leq 2^{n-1}$）より，

– $2^m \leq |z| \leq 2^{d-1}$ となり，構文木の最長路の長さdは必ずm+1以上でなければならない

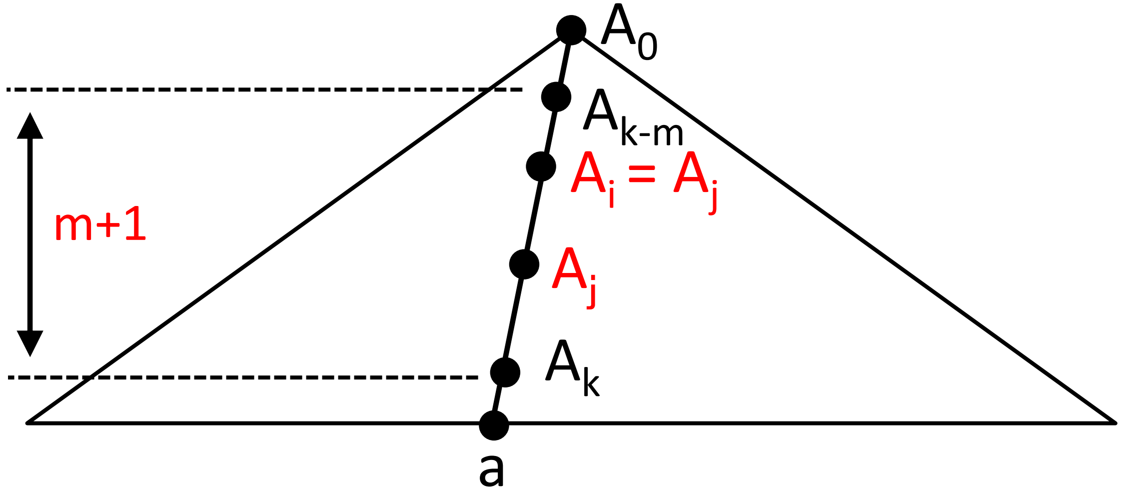
証明：導出中に出現する変数

- 長さm+1以上の経路を任意に選ぶ
  - 経路長： k+1 (≥m+1)
  - 変数A_0, A_1, … , A_k：選んだ経路に出現する変数
    - 根から出現する順序で，同一変数の重複出現を許す
    - Vにはm個の変数しかないので，必ず重複出現が存在



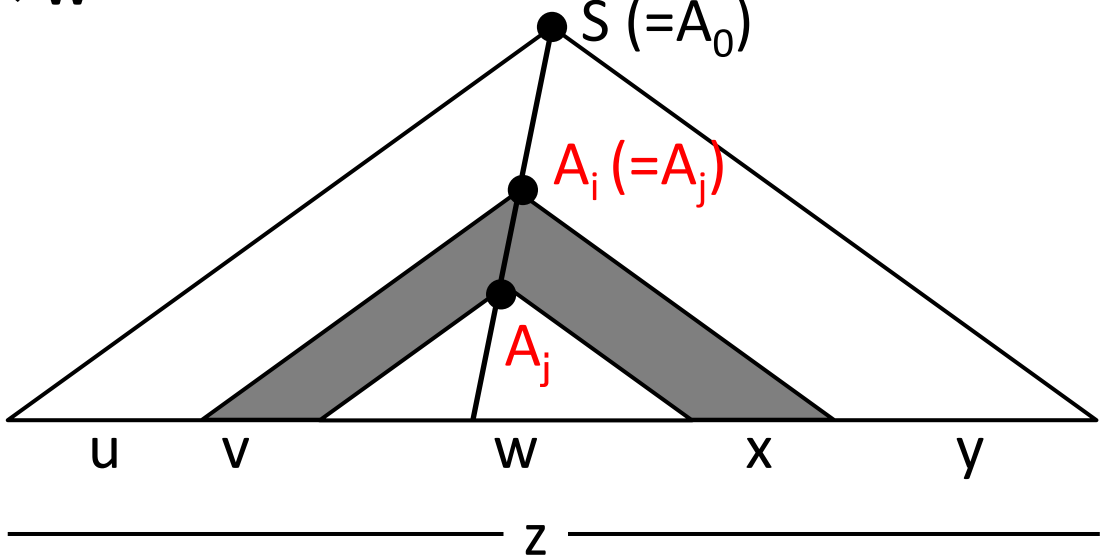
証明：同一変数に着目

- 同一変数A_iとA_jを選ぶ
  - A_{k-m}からA_kまでのm+1個の変数に含まれる同一変数に着目
    - Gの変数の数がm個なので，必ず重複がある
  - 同じ変数が異なる場所に重複して出現
  - このとき，$k-m \le i < j \le k$ が成立（後で使う）



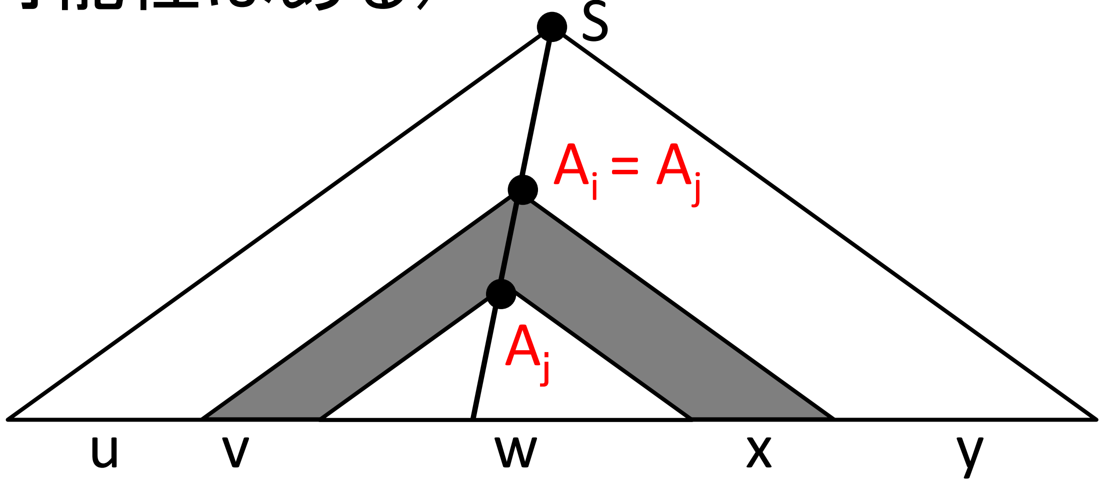
証明：変数と生成列の関係

- A_i , A_jと生成列との関係
  - u, v, w, x, y, zを下図のような終端記号列とすると
  - $S \overset{*}{\Rightarrow} uA_iy \overset{*}{\Rightarrow} uvwxy\ (=z)$
  - $A_i\ (=A_j) \overset{*}{\Rightarrow} vA_jx \overset{*}{\Rightarrow} vwx$
  - $A_j \overset{*}{\Rightarrow} w$



条件2：vx ≠ ε の証明

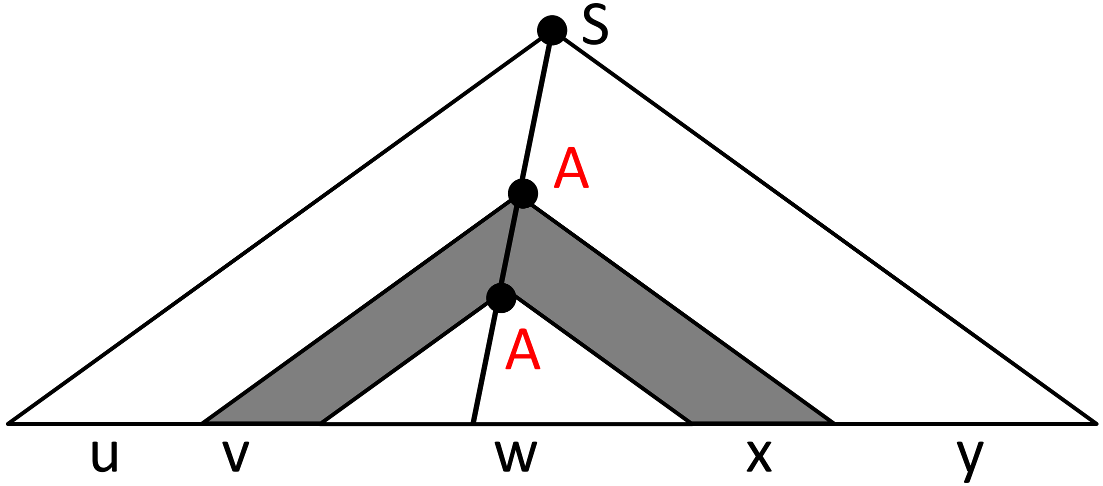
- CNFであるため，生成規則はA_i→BCの形
  - 変数BからA_j（いずれwとなる変数）が導出される場合も，変数CからA_jが導出される場合も，残りの変数から何かしらの終端記号が必ず生成される
  - 従って，vとxがともにεとなることはない（一方がεの可能性はある）



条件2の証明終わり

条件3: $uv^iwx^iy \in L$ の意味

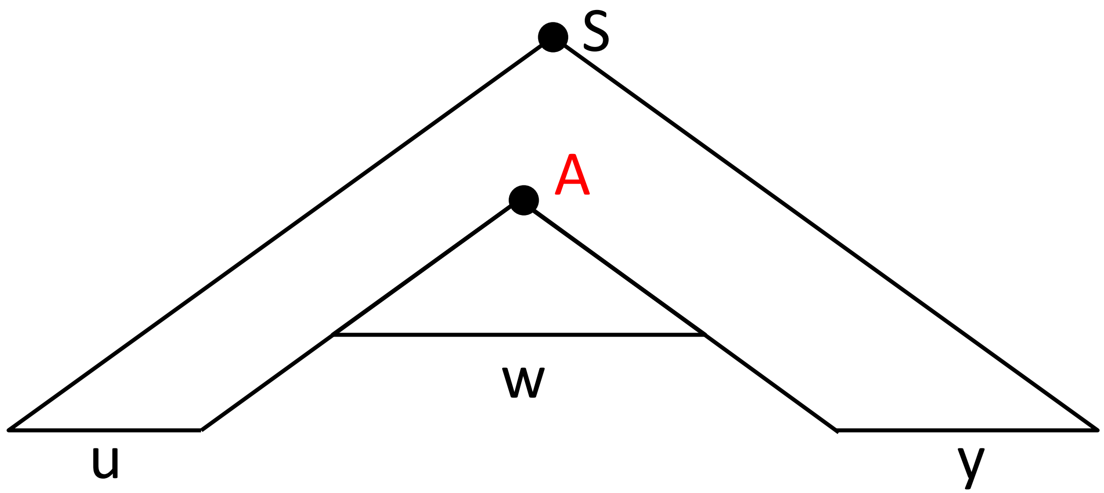
- $A = A_i = A_j$とおく



- uvwxyの導出は
  - $S \overset{*}{\Rightarrow} uAy \overset{*}{\Rightarrow} uvAxy \overset{*}{\Rightarrow} uvwxy\ (=z)$
- つまり, $A \overset{*}{\Rightarrow} vAx$ を何度も繰り返すことが出来る
  - $A \overset{*}{\Rightarrow} w$でもある

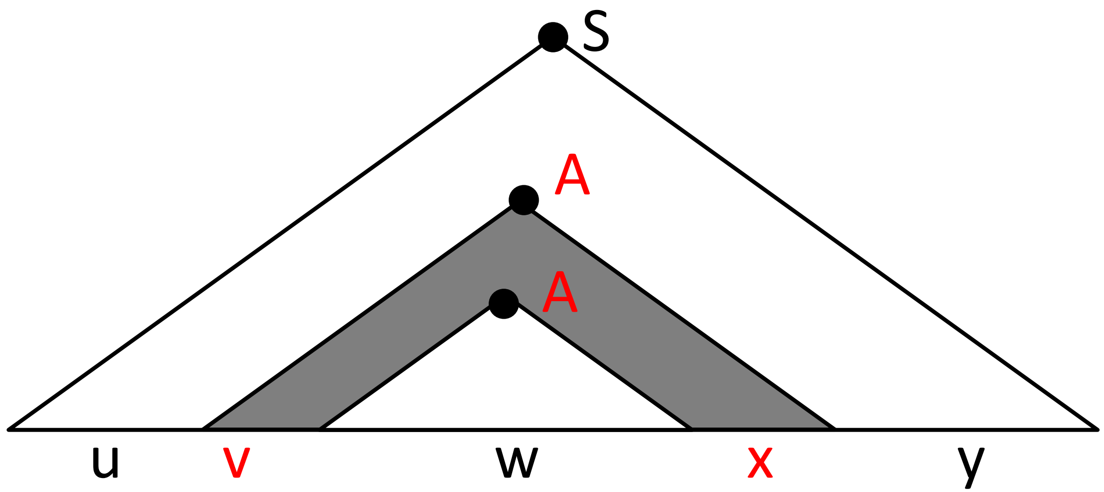
条件3： $uv^iwx^iy \in L$ の証明

- i = 0 のとき
  $S \overset{*}{\Rightarrow} uAy \overset{*}{\Rightarrow} uwy \ (\in L)$



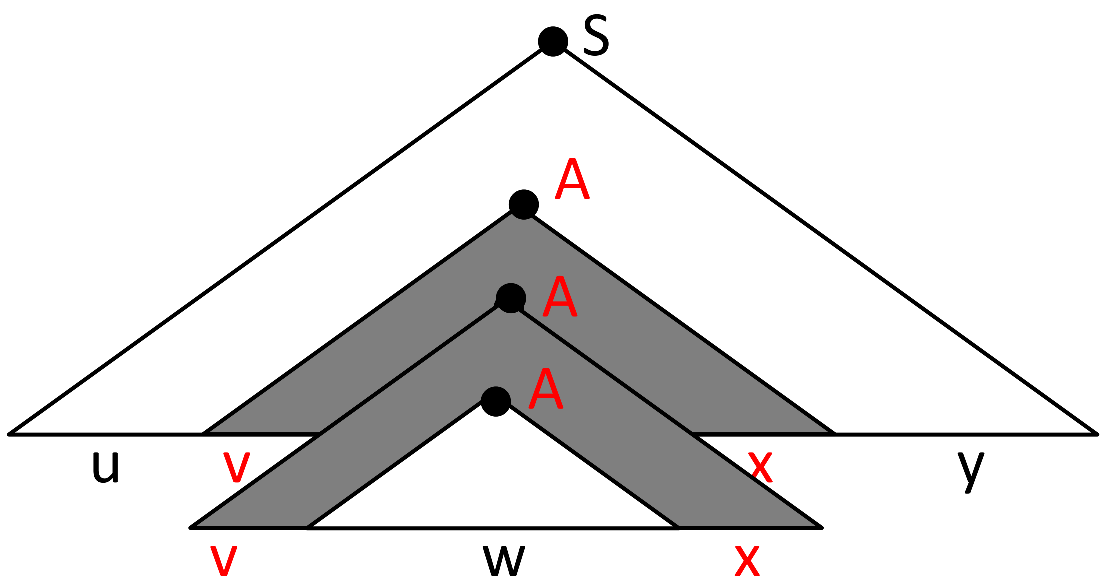
条件3：$uv^iwx^iy \in L$の証明

- $i = 1$ のとき
  $S \overset{*}{\Rightarrow} uAy \overset{*}{\Rightarrow} uvAxy \overset{*}{\Rightarrow} uvwxy\ (\in L)$



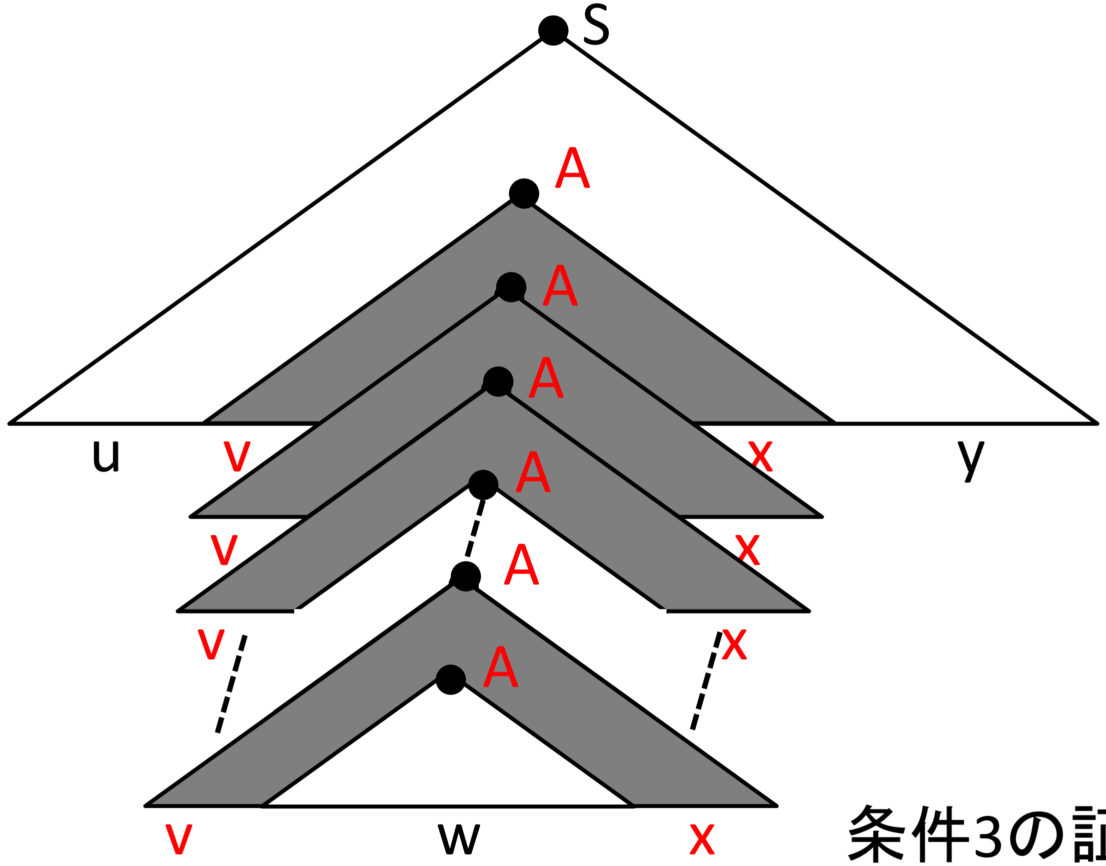
条件3: $uv^iwx^iy \in L$の証明

- $i = 2$ のとき
  $S \overset{*}{\Rightarrow} uAy \overset{*}{\Rightarrow} uvAxy \overset{*}{\Rightarrow} uvvAxxy \overset{*}{\Rightarrow} uvvwxxy\ (\in L)$



条件3：$uv^iwx^iy \in L$の証明

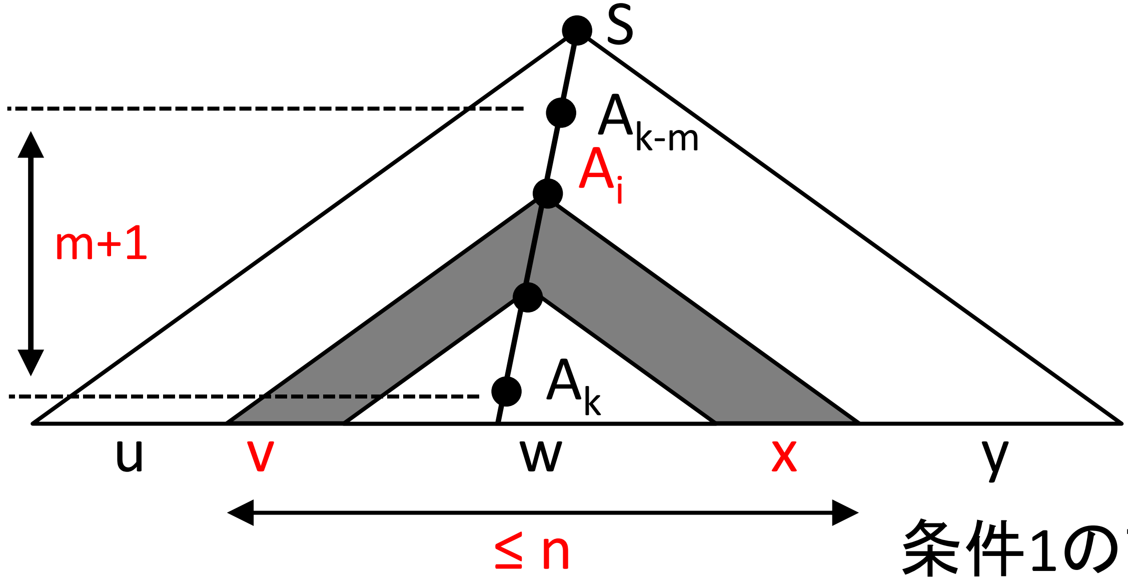
- i 回繰り返したとき
  $S \overset{*}{\Rightarrow} uAy \overset{*}{\Rightarrow} uv^iAx^iy \overset{*}{\Rightarrow} uv^iwx^iy \ (\in L)$



条件3の証明終わり

条件1: |vwx|≤n の証明

- A_i, A_jの選び方より，$k-m \le i < j \le k$
- A_iを根とする部分木の最長路は長さm+1以下
- 定理7.17（$|w| \le 2^{n-1}$）より，A_iから導出されるvwxのサイズは高々$2^{(m+1)-1}=2^m = n$



条件1の証明終わり

7.2.3 反復補題のCFLへの応用

例7.19： $L = \{0^i1^i2^i \mid i \geq 0\}$

反復補題でnを使うので
iで表現しています

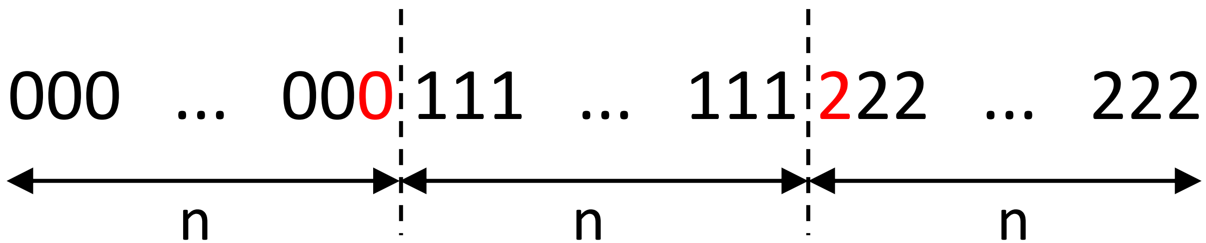
例7.19： $L = \{0^i1^i2^i \mid i \geq 0\}$の解釈

- L : 0,1,2が同数出現　→ LはCFLではない
- 背理法を用いて証明
  - LがCFLと仮定して矛盾を導く
  - n: 反復補題で定める変数
    - nは与えられる数
  - 今回は(与えられたnに対して) $z = 0^n1^n2^n$ を選ぶ

$$\underbrace{000 \quad \ldots \quad 000}_{n}\ \underbrace{111 \quad \ldots \quad 111}_{n}\ \underbrace{222 \quad \ldots \quad 222}_{n}$$

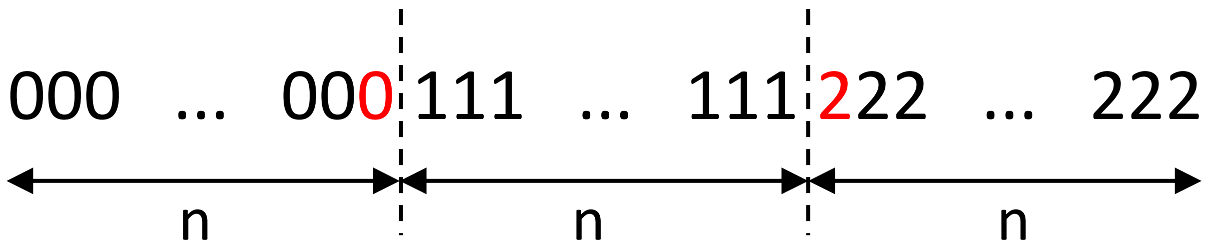
例7.19： $L = \{0^i 1^i 2^i \mid i \geq 0\}$

- $z = uvwxy$ （ただし $|vwx| \leq n$, $vx \neq \varepsilon$）
  - z中の最後の0と最初の2はn+1離れている
    → どうvwxを選んでも0と2をともに含むことはない
    - $|vwx| \leq n$ のため
  - vwxが2を含まない場合と0を含まない場合で場合分け



例7.19：$L = \{0^i1^i2^i \mid i \geq 0\}$

- 2を含まないとき
  - vwxは0と1により構成される
  - $vx \neq \varepsilon$より，vxには0か1が少なくとも1つは含まれる
  - そのとき，0回の反復 $uv^0wx^0y = uwy$は
    - 0か1の少なくとも一方はn個未満
    - 2はn個
- 0を含まないときも同様
- よって，uwyはLに属さず，矛盾



例7.20： $L = \{0^i 1^j 2^i 3^j \mid i \geq 0, j \geq 0\}$

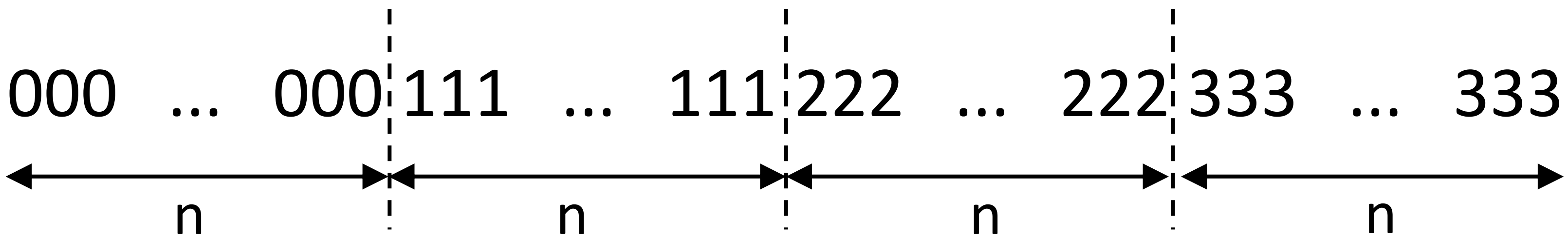
例7.20： L = {$0^i1^j2^i3^j$ | i ≥ 0, j ≥ 0} の解釈

- L : 0と2が同数 かつ 1と3が同数並んだ言語
- 背理法を用いて証明
  - LがCFLと仮定して矛盾を導く
  - n: 反復補題で定める変数
  - $z = 0^n1^n2^n3^n$ を選ぶ

000 … 000	111 … 111	222 … 222	333 … 333
n	n	n	n

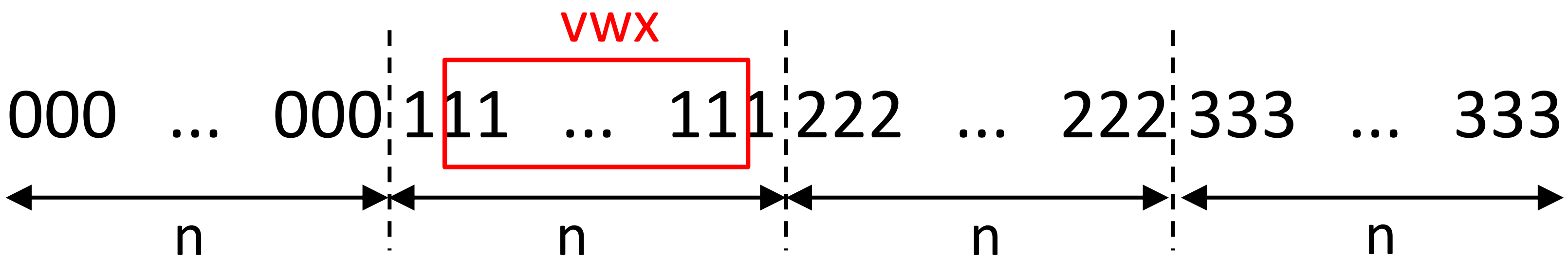
例7.20： $L = \{0^i 1^j 2^i 3^j \mid i \geq 0, j \geq 0\}$

- $z = uvwxy$ （ただし $|vwx| \leq n$, $vx \neq \varepsilon$）
  - $|vwx| \leq n$ なので，vwx中の記号は1種類か，隣接する2種類
  - 出現文字が1種類/2種類で場合わけ



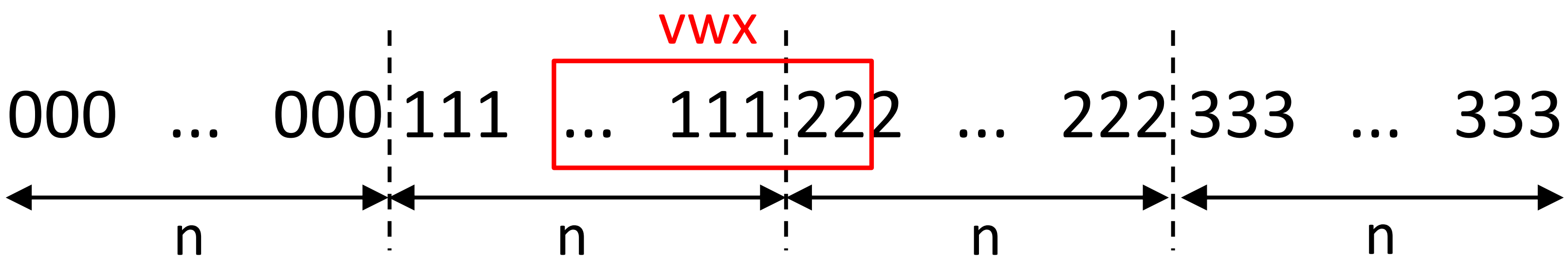
例7.20：$L = \{0^i 1^j 2^i 3^j \mid i \geq 0, j \geq 0\}$

- vwxが1種類の記号のみで構成のとき
  - $vx \neq \varepsilon$より，vxには同記号文字が少なくとも1つは含まれる
  - そのとき，0回の反復 $uv^0wx^0y = uwy$は
    - 該当する記号はn個未満
    - その他の3種類の記号(ペアの記号含む)はn個
  - よって，uwyはLに属さない



例7.20：$L = \{0^i1^j2^i3^j \mid i \geq 0, j \geq 0\}$

- vwxが2種類の記号で構成のとき
  - $vx \neq \varepsilon$より，vxにはその2種類の記号のうちの少なくとも一種の文字が少なくとも1つは含まれる
  - そのとき，0回の反復 $uv^0wx^0y = uwy$は
    - vxに含まれていた記号はn個未満
    - ペアの記号はn個のまま
  - よって，uwyはLに属さない
- 従って，Lは反復補題を満たさず，CFLではない



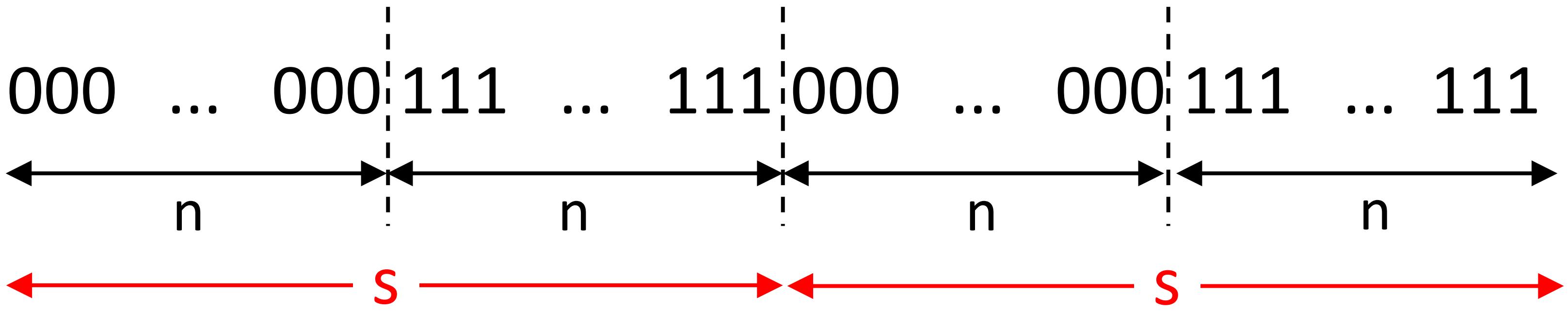
例7.21： $L = \{ss \mid s \in \{0,1\}^*\}$

反復補題でwを使うので

sで表現しています

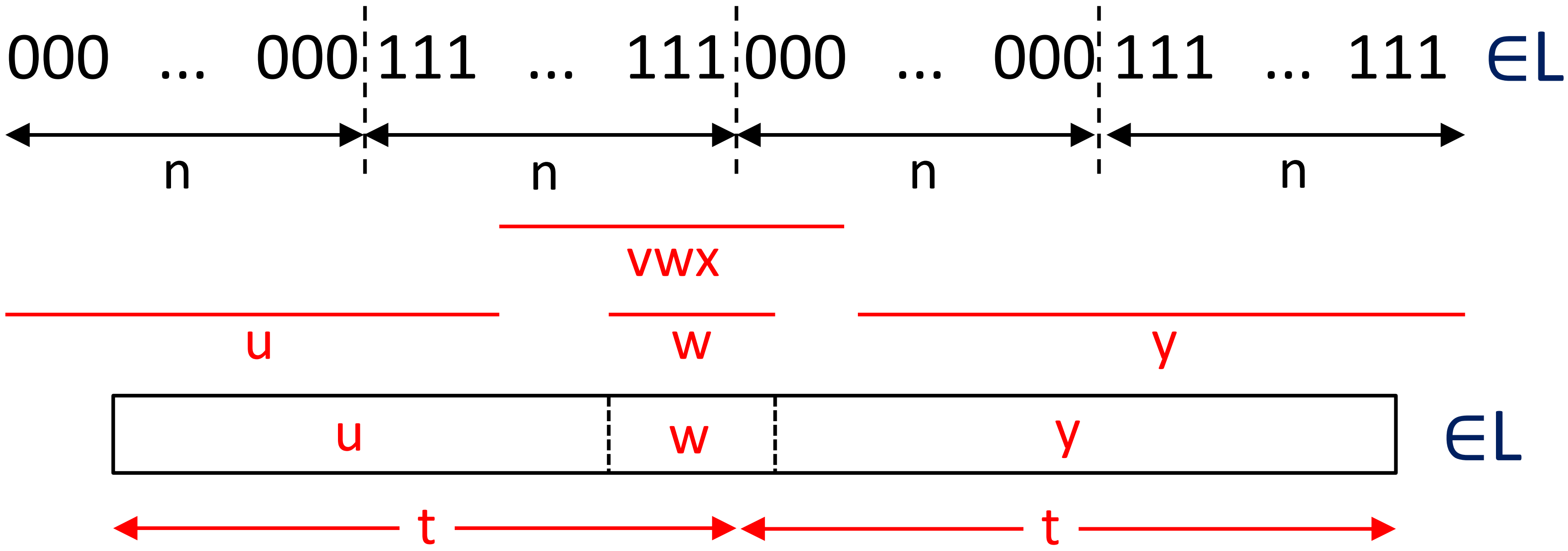
例7.21： L = {ss| s∈{0,1}*}の解釈

- L:2つの同一の記号列が並んだ言語
- 背理法を用いて証明
  - LがCFLと仮定して矛盾を導く
  - n: 反復補題で定める変数
  - $z = 0^n1^n0^n1^n$を選ぶ



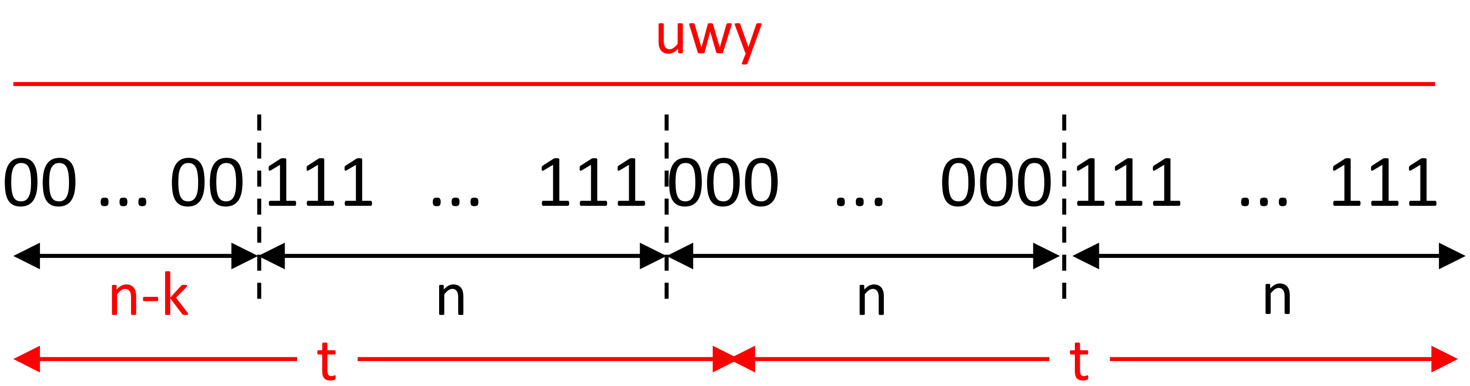
例7.21: $L = \{ss \mid s \in \{0,1\}^*\}$

- $z = uvwxy$ (ただし $|vwx| \le n$, $vx \ne \varepsilon$)
- $|vwx| \le n$ なので，$|uwy| \ge 3n$
  - 0回の繰り返しが$uv^0wx^0y = uwy$で，これもLに属するはずなので，uwyはttの形となるはず



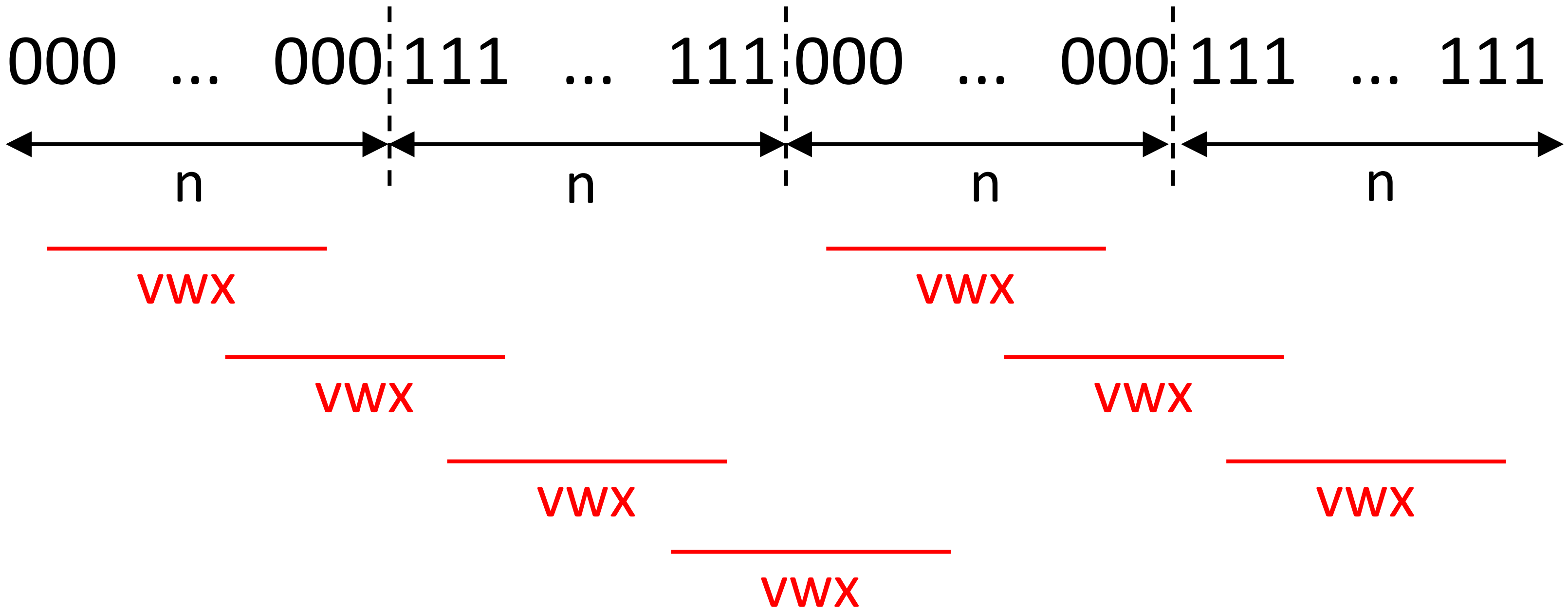
例7.21：$L = \{ss \mid s \in \{0,1\}^*\}$

- vwxの位置で場合わけ
- vwxが最初の0^nに含まれる場合
  - $vx = 0^k$ $(0<k\leq n)$と書ける
  - そのとき，0回の反復 $uv^0wx^0y = uwy$は$0^{n-k}1^n0^n1^n$と書ける
  - これがttだとすると，最初のtの最後の文字は0^nのどこかのはず
  - しかし2つ目のtの最後の文字は1なので，ttの形をしていない．よって，uwyはLに属さない



例7.21：L = {ss | s∈{0,1}*}

- その他のいずれの場合も同様に矛盾を示すことが出来る（テキスト参照）
- よって，Lは反復補題を満たさないため，CFLではない



ミニレポート

ミニレポート:13

- テキストp308 問7.2.1(b)
  - CFLの反復補題を用いて，次の言語が文脈自由でないことを示せ．
  - $L = \{a^n b^n c^i \mid i \leq n\}$

出席レポート

本日の出席レポート

- テキストp308 問7.2.1(a):
  - CFLの反復補題を用いて，次の言語が文脈自由でないことを示せ．
  - $L = \{a^i b^j c^k \mid i < j < k\}$