



アルゴリズムとデータ構造III

2回目: 10月09日

チューリング機械, 文脈自由文法

授業資料 <http://ir.cs.yamanashi.ac.jp/~ysuzuki/algorithm3/index.html>

授業の予定(中間試験まで)

1	10/02	スタック(後置記法で書かれた式の計算)
2	10/09	チューリング機械, 文脈自由文法
3	10/16	構文解析 CYK法
4	10/23	構文解析 CYK法
5	10/30	構文解析 CYK法
6	11/06	構文解析 チャート法
7	11/13	グラフ(動的計画法, ダイクストラ法, DPマッチング)
8	11/20	グラフ(DPマッチング, ビームサーチ, A*アルゴリズム)
9	11/27	中間試験

授業の予定(中間試験以降)

10	12/04	全文検索アルゴリズム (simple search, KMP)
11	12/11	全文検索アルゴリズム (BM, Aho-Corasick)
12	12/18	全文検索アルゴリズム (Aho-Corasick), データ圧縮
13	01/08	暗号 (黄金虫, 踊る人形) 符号化 (モールス信号, Zipfの法則, ハフマン符号) テキスト圧縮
14	01/15	テキスト圧縮 (zip), 音声圧縮 (ADPCM, MP3, CELP), 画像圧縮 (JPEG)
15	01/29	期末試験

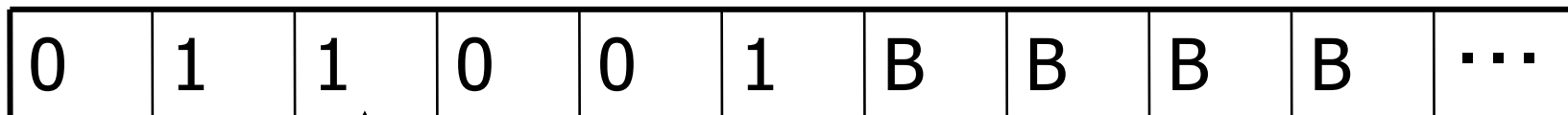
形式言語と有限オートマトン入門

4.5.2 チューリング機械

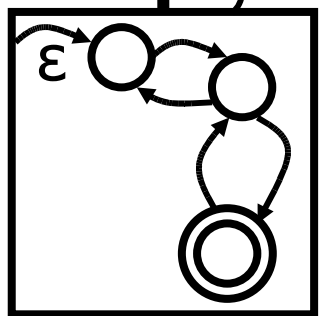
重要!

- 言語受理能力が最も高いオートマトン
- 半無限長の読み書きが自由にできるテープを用いた有限状態機械

読み書きテープ (初期状態では入力語が記述されている)



読み書きヘッド (初期状態: 左端 語の先頭文字位置
テープ上を左右に移動, read, rewrite)



有限状態制御部

最終状態に遷移すると停止して入力語を受理する



チューリング機械(TM)の定義

TM $M=(Q,\Sigma,\Gamma,\delta,S,B,F)$

Q : 内部状態の集合

Σ : 入力アルファベット B を含まない

Γ : テープ記号の集合 ($\Gamma \supset \Sigma$)

B : 空白記号 Γ の要素であるが Σ の要素ではない

δ : 状態遷移関数 $\delta: Q \times \Gamma \rightarrow Q \times \Gamma \times \{R, S, L\}$

R : ヘッドを右に移動, S : ヘッドを移動させない,

L : ヘッドを左に移動

S : 初期状態 $S \in Q$

F : 最終状態(受理状態)の集合 $F \subset Q$



例題4.71 $w_1=0101$

$Q=\{q_0, q_1, q_f\}$, $\Sigma=\{0, 1\}$, $\Gamma=\{0, 1, b\}$, $S=q_0$, $B=b$, $F=\{q_f\}$

δ	0	1	b
q0	(q_0, b, R)	(q_1, b, R)	$(q_f, 0, S)$
q1	(q_1, b, R)	(q_0, b, R)	$(q_f, 1, S)$
qf	---	---	---

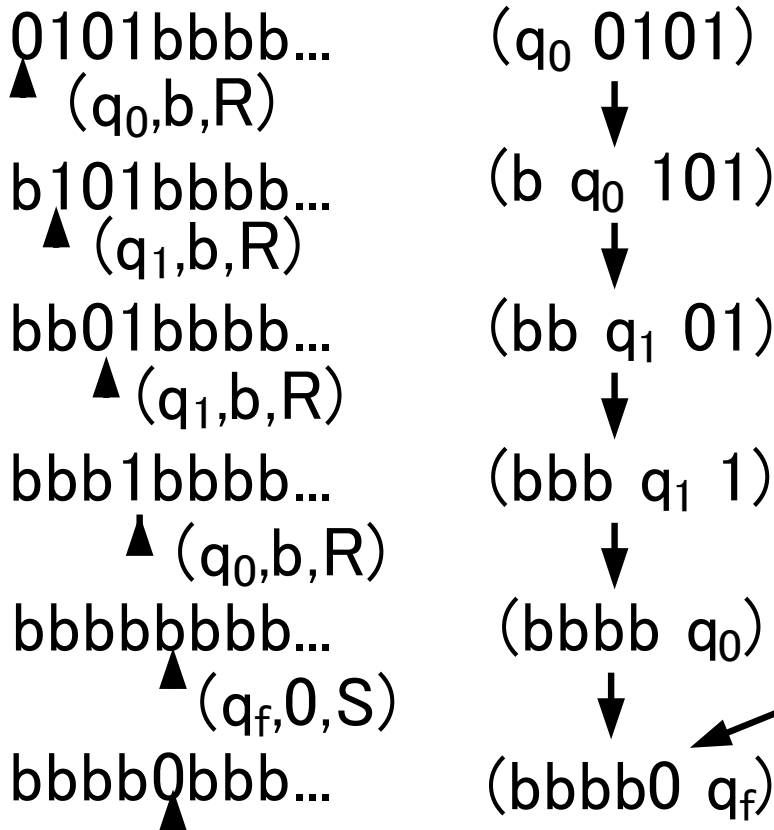
計算状況を示せ.

Σ^* 上の任意の語と, その最終計算状況におけるテープ上の記号との対応を答えよ

例題4.71 答え

$w1=0101$

時点表示(計算状況)



δ	0	1	b
q_0	(q_0, b, R)	(q_1, b, R)	$(q_f, 0, S)$
q_1	(q_1, b, R)	(q_0, b, R)	$(q_f, 1, S)$
q_f	---	---	---

w: 1が奇数個 → 1を出力
 w: 1が偶数個 → 0を出力

最終状態 q_f に遷移 → $w1$ を受理

例題4.71 $w_2' = 011010$

$Q = \{q_0, q_1, q_f\}$, $\Sigma = \{0, 1\}$, $\Gamma = \{0, 1, b\}$, $S = q_0$, $B = b$, $F = \{q_f\}$

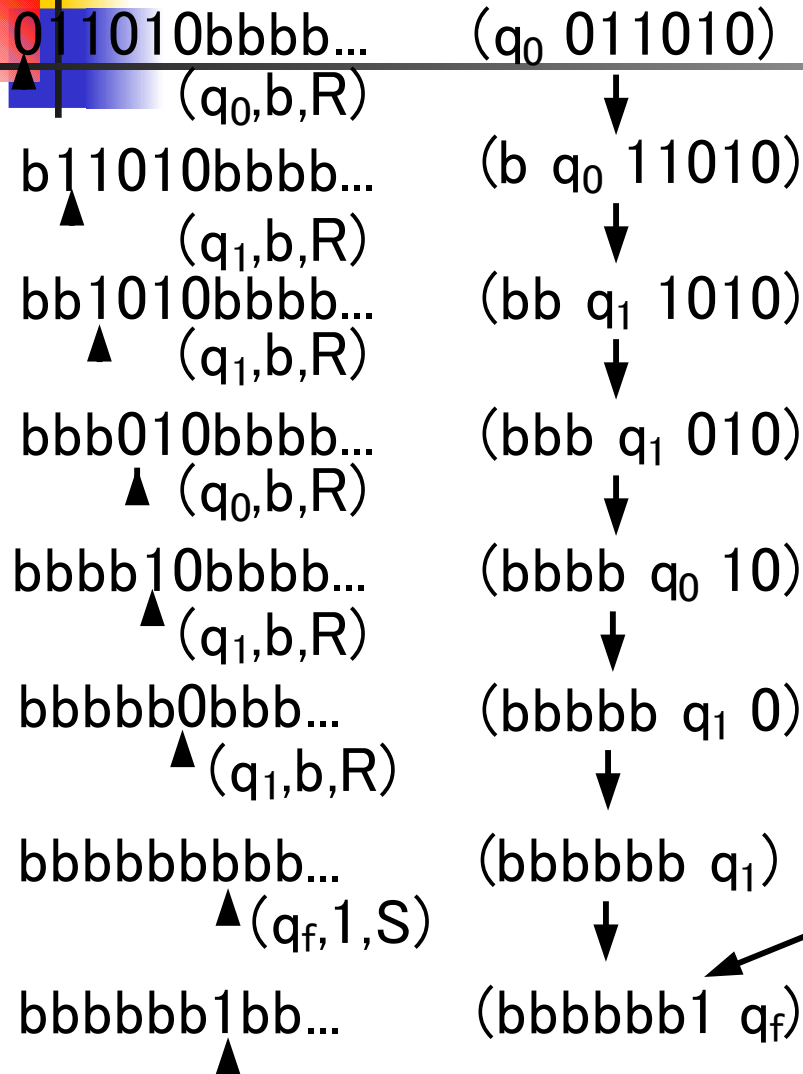
δ	0	1	b
q_0	(q_0, b, R)	(q_1, b, R)	$(q_f, 0, S)$
q_1	(q_1, b, R)	(q_0, b, R)	$(q_f, 1, S)$
q_f	---	---	---

計算状況を示せ.

Σ^* 上の任意の語と, その最終計算状況におけるテープ上の記号との対応を答えよ

例題4.71 答え $W2'=011010$

時点表示(計算状況)



w: 1が奇数個 → 1を出力
w: 1が偶数個 → 0を出力

最終状態q_fに遷移 → w2を受理

練習問題1

例題4.71 $w_2=01101$

$Q=\{q_0, q_1, q_f\}$, $\Sigma=\{0, 1\}$, $\Gamma=\{0, 1, b\}$, $S=q_0$, $B=b$, $F=\{q_f\}$

δ	0	1	b
q0	(q0,b,R)	(q1,b,R)	(qf,0,S)
q1	(q1,b,R)	(q0,b,R)	(qf,1,S)
qf	---	---	---

計算状況を示せ.

Σ^* 上の任意の語と, その最終計算状況におけるテープ上の記号との対応を答えよ

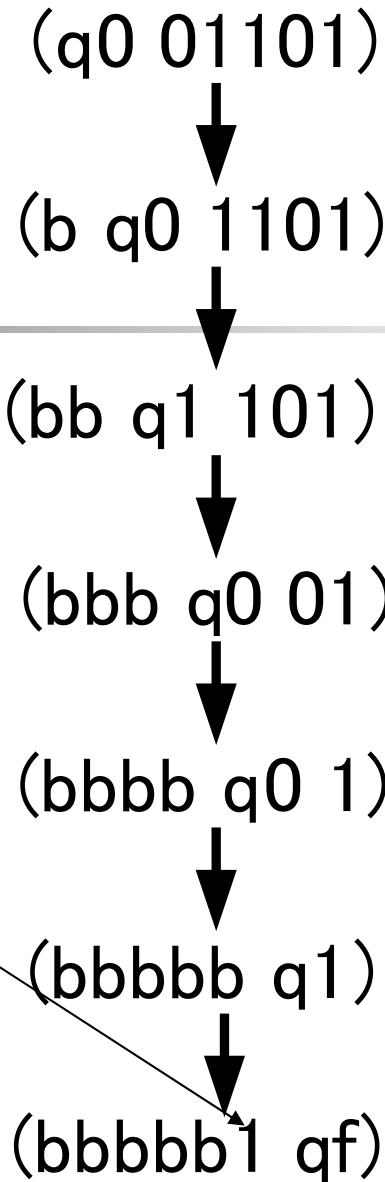
練習問題1

例題4.71 答え

$w_2 = 01101$

w : 1が奇数個 → 1を出力
 w : 1が偶数個 → 0を出力

最終状態 → 受理





4.5.3 オートマトンと計算理論

オートマトンの受理する言語クラス

オートマトン	受理言語型	言語クラス
チューリング機械	第0型言語	句構造言語 (PSL)
線形拘束チューリング機械	第1型言語	文脈依存言語 (CSL)
プッシュダウンオートマトン	第2型言語	文脈自由言語 (CFL)
有限オートマトン	第3型言語	正規言語 (RL)

$RL \subset CFL \subset CSL \subset PSL$ (チョムスキーの言語階層) 12



万能チューリングマシン

- 任意のTMについて, その動作表を与えられるとあたかもそのTMのように振る舞うTM
- コンピュータ
 - プログラム = 動作表 (状態遷移関数表)
 - 入力 = 入力語
 - コンピュータは万能TM
- チューリングテスト
 - TM M が人間
 - コンピュータ(TM) がTM M を完全に模倣できるか

「形式言語と有限オートマトン入門」

5 形式言語理論入門

- 5.1 形式言語理論
- **5.2 文脈自由文法**
- 5.3 線形文法と正規言語
- 5.4 形式言語のクラス階層とオートマトン
- 5.5 言語処理への応用



形式文法Gの定義

- $G=(N,T,P,S)$
 - N: 非終端記号の集合
 - T: 終端記号の集合
 - P: プロダクション
 - S: 開始記号

5.2 文脈自由文法

- 文脈自由文法 (CFG)
 - 文脈自由プロダクションのみから構成される
 - 文脈自由プロダクション
 - $a \rightarrow \beta$
 - ただし, $a \in N$, $\beta \in V^*$
 - N : 非終端記号の集合, T : 終端記号の集合, V : N と T の直和
 - 左辺が変数1つ
- 文脈依存文法 (CSG)
 - 文脈依存プロダクションを含むプロダクションから構成される
 - 文脈依存プロダクション
 - $uav \rightarrow u\beta v$ ただし, $a \in N$, $u, v \in V^*$, $\beta \in V^+$
 - N : 非終端記号の集合, T : 終端記号の集合, V : N と T の直和
 - $u=v=\epsilon$ のとき $(a \rightarrow \beta)$ 文脈自由プロダクションとなる



文脈自由文法の例(例題5.9)

- CFG $G=(N,T,P,S)$
 - N (非終端記号) = $\{B,S\}$
 - T (終端記号) = $\{a,b\}$
 - $P: S \rightarrow aSB \mid ab$
 $B \rightarrow b$
- 語 $aaabbb$ の導出過程
- $L(G)$ はどのような言語か

例題5.9の解答例

■ CFG $G=(N,T,P,S)$

■ $N=\{B,S\}$

■ $T=\{a,b\}$

■ $P: S \rightarrow aSB \mid ab, B \rightarrow b$

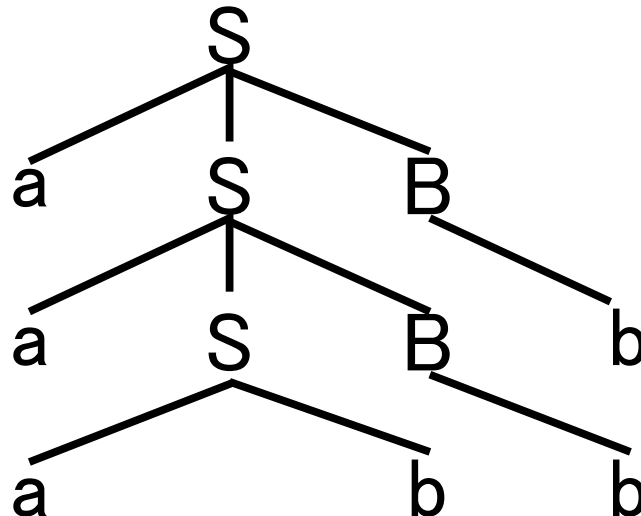
■ $S \Rightarrow aSB \Rightarrow aaSB \Rightarrow aaabBB \Rightarrow aaabbB \Rightarrow aaabbb$

■ $L(G): a^n b^n$

■ 正規表現では表せない

■ プッシュダウンオートマトンでは表現可能

■ 構文木



練習問題2

例題5.10 文脈依存文法の例

- CSG $G=(N,T,P,S)$
- $N=\{A,B,S\}$
- $T=\{a,b\}$
- $P: S \rightarrow aSBA \mid abA, AB \rightarrow BA, bB \rightarrow bb,$
 $bA \rightarrow ba, aA \rightarrow aa$
- 語 $aabbaa$ の導出過程
- $L(G)$ はどのような言語か

練習問題2 解答

例題5.10 aabbaa

- CSG $G=(N,T,P,S)$
- $N=\{A,B,S\}$
- $T=\{a,b\}$
- $P: S \rightarrow aSBA \mid abA, AB \rightarrow BA, bB \rightarrow bb, bA \rightarrow ba, aA \rightarrow aa$

- 語 aabbaa の導出過程

- $S \Rightarrow aSBA \Rightarrow aabABA \Rightarrow aabBAA \Rightarrow aabbAA$

- $\Rightarrow aabbaA \Rightarrow aabbaa$

- $L(G)$ はどのような言語か

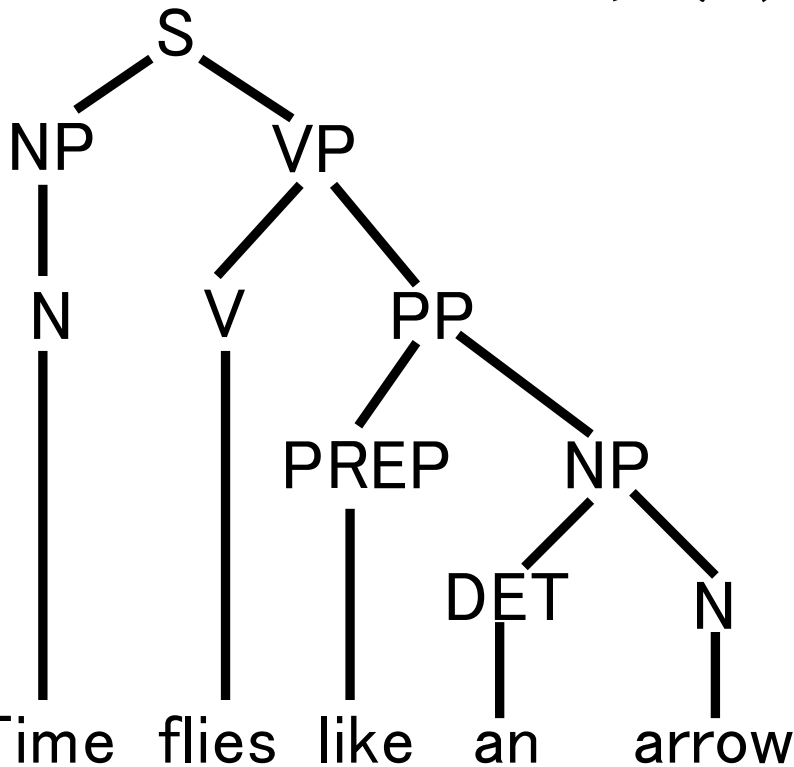
- $L(G): a^n b^n a^n$

構文木(導出木)

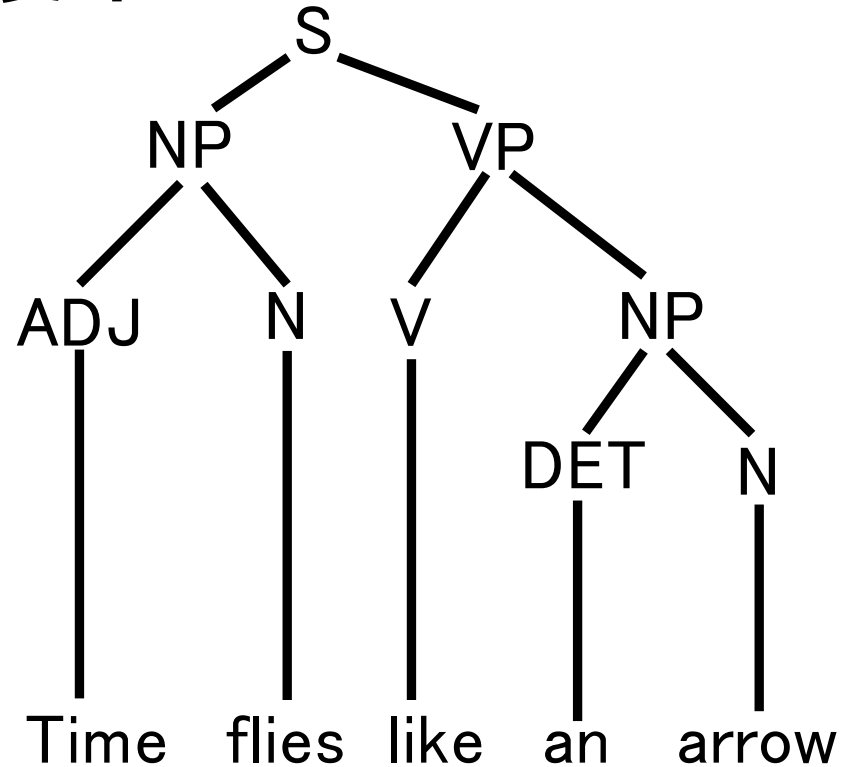
- Time flies like an arrow.

S → NP VP
NP → N | DET N | ADJ N
VP → V PP | V NP
PP → PREP NP
N → Time | arrow | flies
V → flies | like
PREP → like
DET → an

2種類の導出木
→ 文法が曖昧



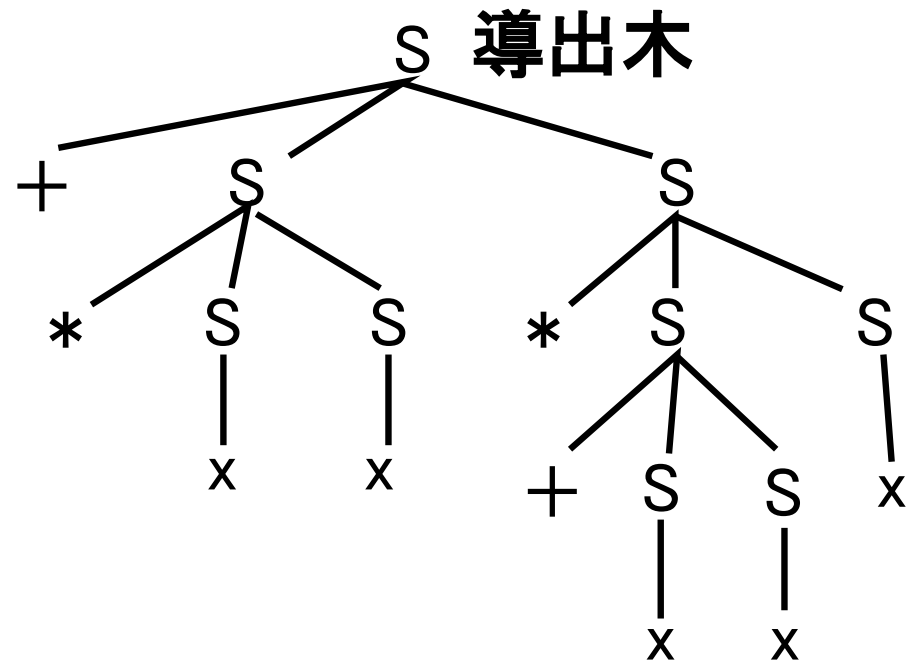
光陰矢の如し



時蠅は矢を好む

例題5.11

- 問題:
- 文法 $N=\{S\}, T=\{x, +, *\}, P=\{S \rightarrow +SS | *SS | x\}$
- 語 $w = +*xx^*+xxx$ を導出せよ
- 語 w の導出木



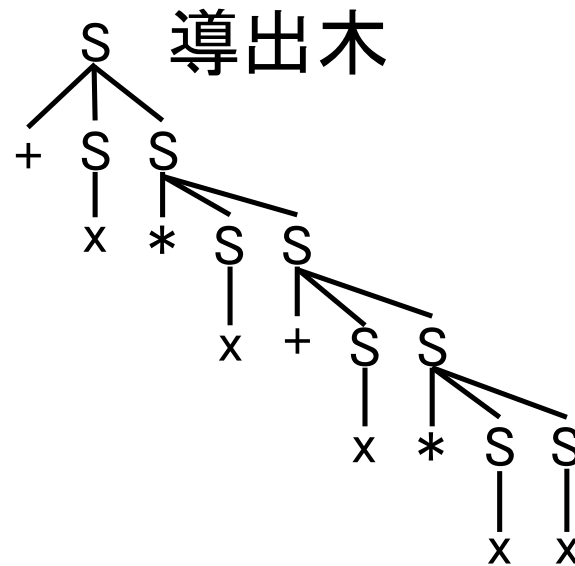
- 解答例

- 導出:

$S \Rightarrow +SS \Rightarrow +*SSS \Rightarrow +*xSS \Rightarrow +*xxS \Rightarrow +*xx^*SS \Rightarrow +*xx^*+S$
 $SS \Rightarrow +*xx^*+xxx$

例題5.12 ①

- 問題
- 文法 $N=\{S\}, T=\{x, +, *\}, P=\{S \rightarrow +SS \mid *SS \mid x\}$
- 中置記法 $x+x*(x+x*x)$



- 解答例
- 前置記法 $+x*x+x*xx$
- $S \Rightarrow +SS \Rightarrow +xS \Rightarrow +x*SS \Rightarrow +x*xS \Rightarrow +x*x+SS$
 $\Rightarrow +x*x+xS \Rightarrow +x*x+x*SS \Rightarrow +x*x+x*xS \Rightarrow +x*x+x*xx$

練習問題3 例題5.12 ②

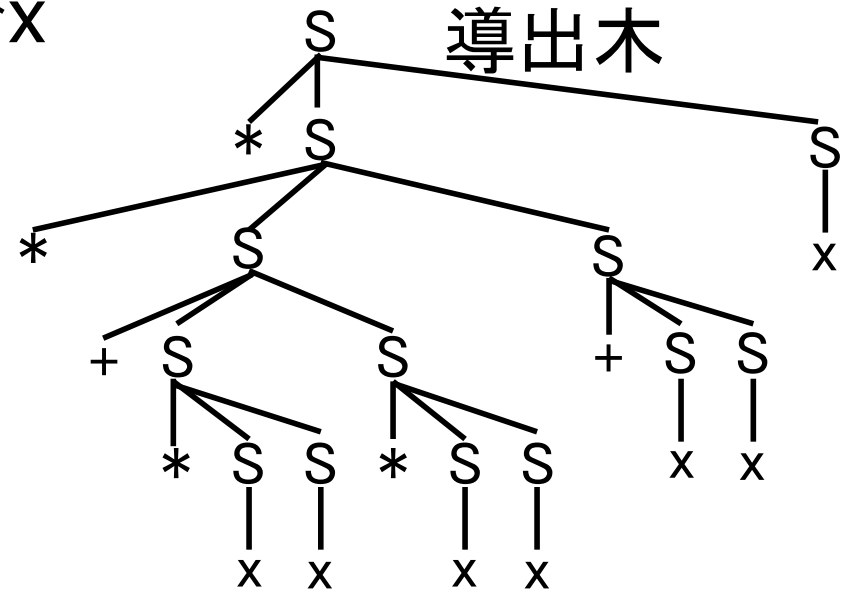
- 問題
- 文法 $N=\{S\}, T=\{x, +, *\}, P=\{S \rightarrow +SS \mid *SS \mid x\}$
- 中置記法 $(x*x+x*x)*(x+x)*x$
- 前置記法
- 最左導出
- 構文木

練習問題3 例題5.12 ②の解答例

■ 問題

■ 文法 $N=\{S\}, T=\{x, +, *\}, P=\{S \rightarrow +SS | *SS | x\}$

■ 中置記法 $(x*x+x*x)*(x+x)*x$



■ 解答例

■ 前置記法 $**+*xx*xx+xxx$

■ $S \Rightarrow *SS \Rightarrow **SSS \Rightarrow **+SSSS \Rightarrow **+*SSSSS \Rightarrow **+*xSSSS$
 $\Rightarrow **+*xxSSS \Rightarrow **+*xx*SSSS \Rightarrow **+*xx*xSSS$
 $\Rightarrow **+*xx*xxSS \Rightarrow **+*xx*xx+SSS$

■ $\Rightarrow **+*xx*xx+xxx$



文脈自由文法の曖昧性

- どのような導出を行っても同じ導出木が得られる
- ⇒ 文法Gは曖昧でない

- 複数の異なった導出木が構成できるような語を含む
- ⇒ 文法Gは曖昧である



例題5.26

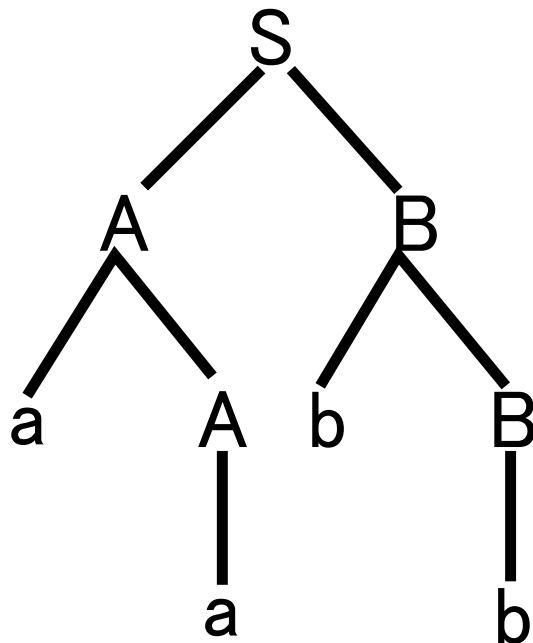
- 文法 $G=(N,T,P,S)$ において, $N=\{S,A,B\}, T=\{a,b\}$,
- $P: S \rightarrow AB \mid aAB, A \rightarrow aA \mid a, B \rightarrow bB \mid b$
- この文法が曖昧であることを示せ

例題5.26 解答例

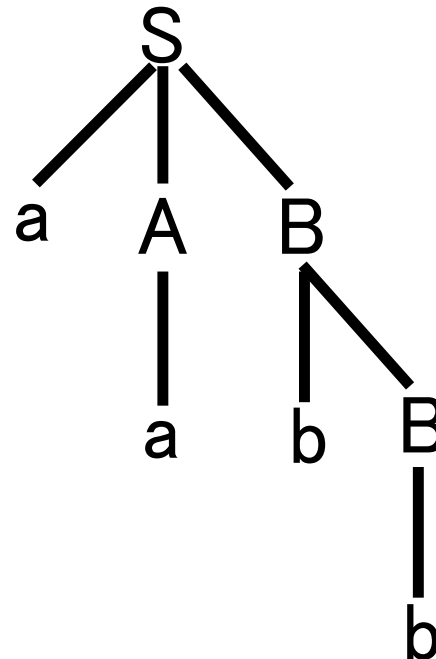
同一文字列に対して2種類の導出
木が構成可能→曖昧である

- 1. $S \rightarrow AB \rightarrow aAB \rightarrow aAbB \rightarrow aabB \rightarrow aabb$
- 2. $S \rightarrow aAB \rightarrow aaB \rightarrow aabB \rightarrow aabb$

1.



2.



練習問題4

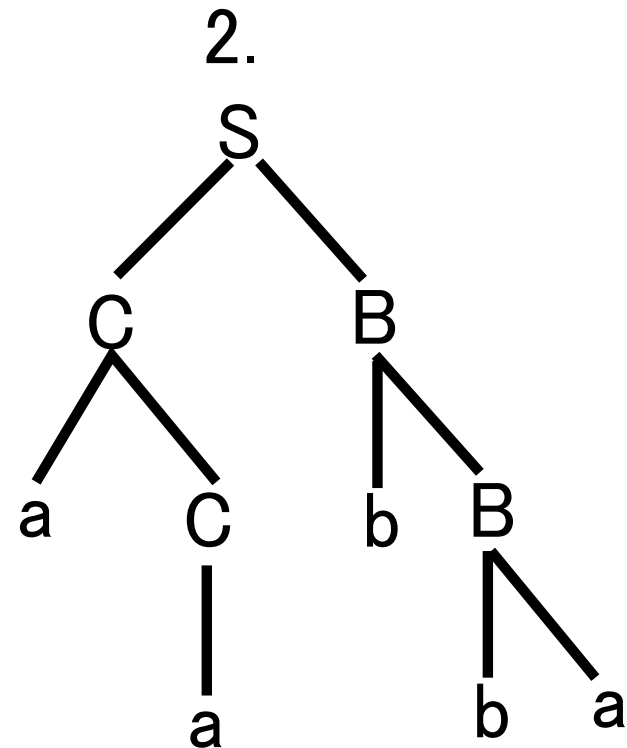
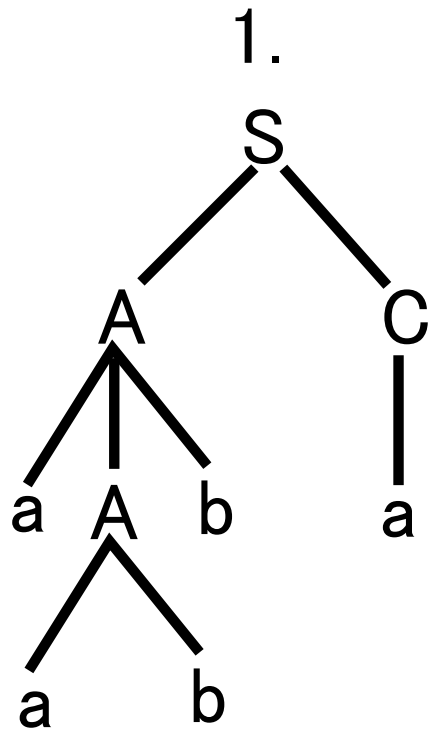
例題5.27

- 文法 $G=(N,T,P,S)$ において,
 - $N=\{S,A,B,C\}, T=\{a,b\},$
 - $P: S \rightarrow AC|CB, A \rightarrow aA|a, A \rightarrow aAb|ab, B \rightarrow bB|ba$
 - $C \rightarrow aC|a$
-
- この文法が曖昧であることを, $aabba$ の導出木を構成して示せ

練習問題4 例題5.27 解答例

同一文字列に対して2種類の導出木が構成可能 → 曖昧である

- 1. $S \rightarrow AC \rightarrow aAbC \rightarrow aAba \rightarrow aabba$
- 2. $S \rightarrow CB \rightarrow aCB \rightarrow aCbB \rightarrow aabB \rightarrow aabba$



CFGの構文図式

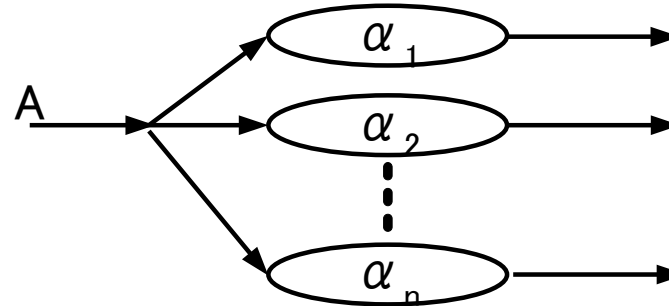
文脈自由プロダクション

$$A \rightarrow \alpha$$

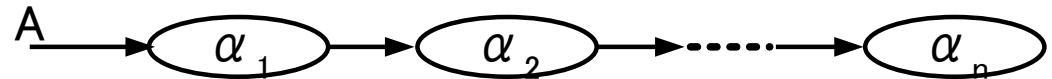
構文図式



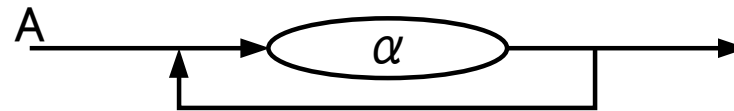
$$A \rightarrow \alpha_1 | \alpha_2 | \dots | \alpha_n$$



$$A \rightarrow \alpha_1 \alpha_2 \dots \alpha_n$$



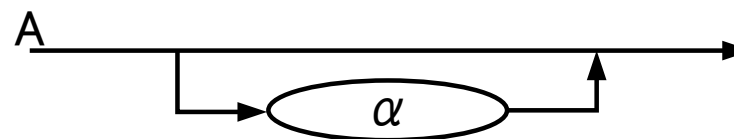
$$A \rightarrow \alpha | \alpha A |$$



$$A \rightarrow \varepsilon | \alpha | \alpha A |$$



$$A \rightarrow \varepsilon | \alpha$$





文脈自由文法の簡単化

- 以下の書き換え規則を削除する
- 詳しくは154ページ
 - 開始記号Sから導出に使われることの無い非終端記号
 - ε -規則 ($A \rightarrow \varepsilon: A \in N$)
 - 単位規則 ($A \rightarrow B: A, B \in N$)

例題5.31 ① ε -規則の消去

■ $S \rightarrow aA, A \rightarrow bA | \varepsilon$

■ $S \rightarrow aA$

■ $A \rightarrow bA$

■ $A \rightarrow \varepsilon$

■ $S \rightarrow aA | a$

■ $A \rightarrow bA | b$

■ $S \rightarrow aA | a, A \rightarrow bA | b$

例題5.31 ② ϵ -規則の消去

■ $S \rightarrow aA, A \rightarrow aA | bB, B \rightarrow bB | \epsilon$

■ $S \rightarrow aA$

■ $A \rightarrow aA$

■ $A \rightarrow bB$

■ $B \rightarrow bB$

■ $B \rightarrow \epsilon$

■ $S \rightarrow aA$

■ $A \rightarrow aA$

■ $A \rightarrow bB | b$

■ $B \rightarrow bB | b$

■ $S \rightarrow aA, A \rightarrow aA | bB | b, B \rightarrow bB | b$

例題5.31 ③ ϵ -規則の消去

■ $S \rightarrow aAB, A \rightarrow aA|a|Bb, B \rightarrow bB|\epsilon$

■ $S \rightarrow aAB$

■ $A \rightarrow aA$

■ $A \rightarrow a$

■ $A \rightarrow Bb$

■ $B \rightarrow bB$

■ $B \rightarrow \epsilon$

■ $S \rightarrow aAB$

■ $A \rightarrow aA$

■ $A \rightarrow a$

■ $A \rightarrow Bb|b$

■ $B \rightarrow bB|b$

■ $S \rightarrow aAB, A \rightarrow aA|a|Bb|b, B \rightarrow bB|b$



例題5.32 ①

- $S \rightarrow aA, A \rightarrow aB|B, B \rightarrow bB|b$



例題5.33 ①

- $S \rightarrow AB|a, A \rightarrow a$
 - $S \rightarrow AB$
 - $S \rightarrow a$
 - $A \rightarrow a$

- $B \rightarrow$ が無いので $S \rightarrow AB$ を削除
- $S \rightarrow a$



文脈自由文法の標準形

- チョムスキー標準形

- 文脈自由文法の規約化された生成規則が,
- すべて $A, B, C \in N$, $a \in T$ として,
- $A \rightarrow BC$ または $A \rightarrow a$

の形をしているとき, この生成規則をチョムスキー標準形という

文脈自由な生成規則のチョムスキー標準形への変換

- $X, A, B, C \in N, a \in T$ として,
- $X \rightarrow aB$ ならば $X \rightarrow AB, A \rightarrow a$ と分解する
- $X \rightarrow ABC$ ならば $X \rightarrow AY, Y \rightarrow BC$ と分解する



例題5.34

- 文法 $G=(N,T,P,S)$ において, $N=\{S,A,B,C\}$, $T=\{a,b\}$, P を, $S \rightarrow AaC \mid CbBa$, $A \rightarrow aAb \mid ab$, $B \rightarrow bB \mid b$, $C \rightarrow Ca \mid a$ とする. この文法 G を
- チョムスキー形生成規則をもつ文脈自由文法に書き換えよ.



例題5.34 解答例

- $S \rightarrow AaC \Rightarrow S \rightarrow AS_1, S_1 \rightarrow S_2C, S_2 \rightarrow a$
- $S \rightarrow CbBa \Rightarrow S \rightarrow CS_3, S_3 \rightarrow S_4S_5, S_4 \rightarrow b, S_5 \rightarrow BS_2$
- $A \rightarrow aAb \Rightarrow A \rightarrow S_2A_1, A_1 \rightarrow AS_4$
- $A \rightarrow ab \Rightarrow A \rightarrow S_2S_4$
- $B \rightarrow bB \Rightarrow B \rightarrow S_4B$
- $C \rightarrow Ca \Rightarrow C \rightarrow CS_2$

次回の講義

構文解析アルゴリズム

- **CYK (Cocke-Younger-Kasami) 法**
 - チョムスキー標準形で書かれた言語の構文解析手法
- チャート法, LR法