文脈自由文法

離散数学・オートマトン
2024 年後期
佐賀大学理工学部　只木進一

1. 形式文法: Formal grammars
2. 正規文法: Regular Grammars
3. 文脈自由文法: Context Free Grammars
4. 文脈自由言語を受理する NPDA: NPDA Accepting CFL
5. 空スタックで受理する NPDA に対応する文脈自由文法: CFL corresponding to NPDA

言語と文法: Languages and grammars

- 言語の構成要素: Elements of languages
  - 文法: Grammars
  - 語: Words
  - 文: Sentences
- 文法: Grammars
  - 語の配置規則: Rules for arranging words
  - 文の生成規則: Rules for generating sentences
  - 生成文法: Generative Grammars
    N. Chomsky, *Syntactic Structure*, 1957.
- 有限オートマトン、プッシュダウンオートマトンの受理言語を記述する文法とは: Grammars describing languages accepted by finite automata and pushdown automata?

形式文法: Formal Grammar

文法の一般的定式化: General formalization of grammars

$$G = \langle N, \Sigma, P, S_0 \rangle \tag{1.1}$$

- N: **非終端アルファベット**: **文法の要素（品詞など）に相当**: Non-terminal alphabet
- Σ: **終端アルファベット**: **語に相当**: Terminal alphabet
- P: **生成規則**: Productions
- $S_0 \in N$: **開始記号**: Start symbol

生成文法 (Generative Grammars) **とも言う**

- **開始記号から終端記号の列を生成**: Generate a sequence of terminal symbols from the start symbol

正規文法: Regular Grammars

- 正規表現に対応した正規言語を生成: Generate regular languages corresponding to regular expressions
- 生成規則: Productions

$$P : N \rightarrow \Sigma N | \Sigma \tag{2.1}$$

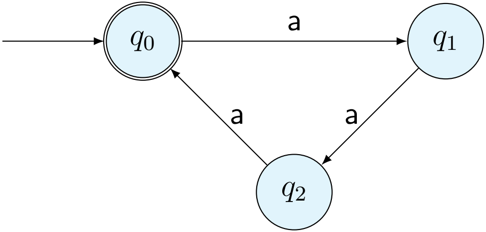
  - ただし、必要ならば $S_0 \rightarrow \epsilon$ も許容する: $S_0 \rightarrow \epsilon$ is allowed if necessary
  - 非終端記号が、一つの終端記号、または一つの終端記号と一つの非終端記号の列に置き換わる
- 記号 | は、`or` を表す: | represents `or`
- 右辺の ΣN が無限回の繰り返しを可能とする: ΣN on the right side allows infinite repetition

$$A \rightarrow \mathsf{a}A \mid \mathsf{a} \tag{2.2}$$

例 2.1:

$$N = \{S_0, S_1, S_2\}, \quad \Sigma = \{\mathsf{a}\},$$
$$P = \{S_0 \to \epsilon | \mathsf{a}S_1, S_1 \to \mathsf{a}S_2, S_2 \to \mathsf{a}S_0 | \mathsf{a}\}$$

$$\begin{aligned} S_0 &\Rightarrow \mathsf{a}S_1 \Rightarrow \mathsf{aa}S_2 \Rightarrow \mathsf{aaa}S_0 \\ &\Rightarrow \mathsf{aaaa}S_1 \Rightarrow \mathsf{aaaaa}S_2 \Rightarrow \mathsf{aaaaaa} \end{aligned}$$



$$L = \left\{\mathsf{a}^{3n} | n \in N \cup \{0\}\right\} = (\mathbf{aaa})^*$$

例 2.2:

$$N = \{S, A\}, \Sigma = \{\mathsf{a}, \mathsf{b}\}$$
$$P = \{S \to \epsilon \mid \mathsf{a}A, A \to \mathsf{b}S\}$$

正規文法が正規表現と同等であること: Regular grammars are equivalent to regular expressions

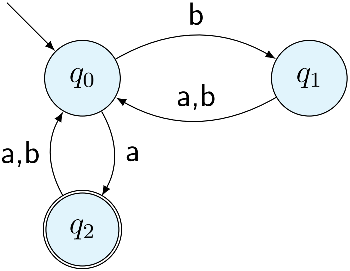
- 決定性有限オートマトンを正規文法に翻訳できること: DFAs can be translated to regular grammars
    - 遷移関数を生成規則に翻訳
- 正規文法を非決定性有限オートマトンに翻訳できること: Regular grammars can be translated to NFAs
    - 生成規則を遷移関数に翻訳
- 一般的な翻訳規則があることが重要: Existence of general translation rules is important

決定性有限オートマトンから正規文法へ: From DFA to regular grammar

Algorithm 1 $M = \langle Q, \Sigma, \delta, q_0, F \rangle \Rightarrow G = \langle N, \Sigma, P, q_0 \rangle$

```
N = Q                                  ▷ 内部状態を非終端記号に読み替える
for all q' = δ(q, a) do                ▷ 遷移関数の各場合を変換
    P に q → aq' を追加
    if q' ∈ F then                     ▷ 終状態への遷移の場合は、終端記号のみへ
        P に q → a を追加
    end if
end for
```

例 2.3:



- **正規表現**: regular expression

$$((\mathrm{a}+\mathrm{b})(\mathrm{a}+\mathrm{b}))^{*}\,\mathrm{a}$$

- **対応する正規文法の非終端記号と終端記号**: Non-terminal and terminal symbols of the corresponding regular grammar

$$N = \{q_0, q_2, q_1\}$$

$$\Sigma = \{\mathrm{a}, \mathrm{b}\}$$

遷移関数	生成規則
$\delta(q_0, \mathsf{a}) = q_2$	$q_0 \to \mathsf{a}q_2 \mid \mathsf{a}$
$\delta(q_0, \mathsf{b}) = q_1$	$q_0 \to \mathsf{b}q_1$
$\delta(q_1, \mathsf{a}) = q_0$	$q_1 \to \mathsf{a}q_0$
$\delta(q_1, \mathsf{b}) = q_0$	$q_1 \to \mathsf{b}q_0$
$\delta(q_2, \mathsf{a}) = q_0$	$q_2 \to \mathsf{a}q_0$
$\delta(q_2, \mathsf{b}) = q_0$	$q_2 \to \mathsf{b}q_0$

以上まとめて

$$P = \{q_0 \to \mathsf{a}q_2 \mid \mathsf{b}q_1 \mid \mathsf{a},\ q_1 \to \mathsf{a}q_0 \mid \mathsf{b}q_0,\ q_2 \to \mathsf{a}q_0 \mid \mathsf{b}q_0\}$$

導出例: Derivation example

$$\begin{aligned} q_0 &\Rightarrow \mathsf{a}q_2 \Rightarrow \mathsf{ab}q_0 \Rightarrow \mathsf{abb}q_1 \\ &= \mathsf{abba}q_0 \Rightarrow \mathsf{abbaa} \end{aligned}$$

正規文法から非決定性有限オートマトンへ: From regular grammar to NFA

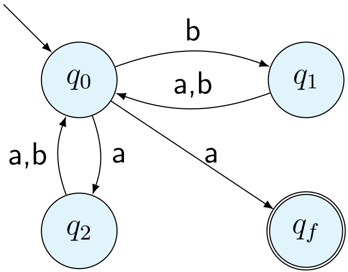
Algorithm 2 $G = \langle N, \Sigma, P, S_0 \rangle \Rightarrow M = \langle Q, \Sigma, \delta, S_0, \rangle$

$Q = N$ ▷ 非終端記号を内部状態に読み替える
$F = \{q_f\}$
for all $A \to aB$ **do** ▷ 非終端記号を含む場合
  $B \in \delta(A, a)$
end for
for all $A \to a$ **do** ▷ 終端記号のみの場合は終状態へ
  $q_f \in \delta(A, a)$
end for
if $S_0 \to \epsilon \in P$ **then**
  $q_f \in \delta(S_0.\epsilon)$
end if

例 2.4:

$$
\begin{aligned}
N &= \{q_0, q_2, q_1\} \\
\Sigma &= \{\mathsf{a}, \mathsf{b}\} \\
P &= \{q_0 \to \mathsf{a}q_2|\mathsf{a}|\mathsf{b}q_1, q_1 \to \mathsf{a}q_0|\mathsf{b}q_0, q_2 \to \mathsf{a}q_0|\mathsf{b}q_0\}
\end{aligned}
$$

CFG	NFA
$q_0 \rightarrow \mathsf{a}q_2 \vert \mathsf{a} \vert \mathsf{b}q_1$	$\delta(q_0, \mathsf{a}) = \{q1, q_f\}$ $\delta(q_0, \mathsf{b}) = \{q2\}$
$q_1 \rightarrow \mathsf{a}q_0 \vert \mathsf{b}q_0$	$\delta(q_1, \mathsf{a}) = \{q_0\}$ $\delta(q_1, \mathsf{b}) = \{q_0\}$
$q_2 \rightarrow \mathsf{a}q_0 \vert \mathsf{b}q_0$	$\delta(q_2, \mathsf{a}) = \{q_0\}$ $\delta(q_2, \mathsf{b}) = \{q_0\}$



文脈自由文法: Context Free Grammars

- 生成規則: 非終端記号が非終端記号または終端記号の長さ 0 以上の列へと変換: Non-terminal symbols are replaced by a sequence of terminal and non-terminal symbols

$$P : N \to (\Sigma \cup N)^* \tag{3.1}$$

- 例: example

$$
\begin{aligned}
N &= \{S_0\}, \ \Sigma = \{\mathsf{a}, \mathsf{b}\} \\
P &= \{S_0 \to \epsilon | \mathsf{a} S_0 \mathsf{b}\} \\
&\quad S_0 \Rightarrow \mathsf{a}S_0\mathsf{b} \Rightarrow \mathsf{aa}S_0\mathsf{bb} \Rightarrow \mathsf{aaa}S_0\mathsf{bbb} \Rightarrow \mathsf{aaabbb}
\end{aligned}
$$

なぜ「文脈自由」なのか: Why they are called *context free*

- **生成規則の左辺は、非終端記号一つ**: A single non-terminal symbol on the left side of a production rule
- **前後の非終端記号や終端記号との繋がり（文脈）を無視**: Ignoring the context of the surrounding symbols

二つの標準形: Standard forms

- チョムスキー標準形 (Chomsky normal form, CNF)
  - $A \to BC (B, C \in N)$
  - $A \to a (a \in \Sigma)$
  - $S \to \epsilon$ も可
- グライバッハ標準形 (Greibach normal form, GNF)
  - $A \to a\alpha\ (a \in \Sigma, \alpha \in N^*)$
  - $S \to \epsilon$ も可

PDA の 3 種類の受理: Three types of acceptance by a PDA

- 入力終了時にスタックが空: Empty stack at the end of input

$$L_A(M) = \{w \in \Sigma * \mid (q_0, w, Z) \vdash^* (q, \epsilon, \epsilon)\} \tag{4.1}$$

- **入力終了時に終状態**: Reaching the final state at the end of input

$$L_A(M) = \{w \in \Sigma * \mid (q_0, w, Z) \vdash^* (q, \epsilon, \gamma), q \in F\} \tag{4.2}$$

- **入力終了時にスタックが空、かつ終状態**: Empty stack and reaching the final state at the end of input

$$L_A(M) = \{w \in \Sigma * \mid (q_0, w, Z) \vdash^* (q, \epsilon, \epsilon), q \in F\} \tag{4.3}$$

文脈自由言語 L を受理する NPDA: NPDA Accepting CFL L

$$G = \langle N, \Sigma, P, S \rangle \tag{4.4}$$

- $L \cap \{\epsilon\} = \emptyset$: **明示的に ϵ を受理する場合を除く**: Except when ϵ is explicitly accepted
- Greibach **標準形**: GNF
- **最左導出 (一番左の非終端記号から生成規則を適用)**: leftmost derivation
- **等価な** NPDA: **入力終了時にスタックが空になる**: Accepting by the empty stack at the end of input

$$M = \langle \{q\}, \Sigma, N, \delta, q, S, \emptyset \rangle \tag{4.5}$$

- **最左導出**: Leftmost derivation

$$S \Rightarrow a_1 A_1 \gamma_1 \Rightarrow a_1 a_2 A_2 \gamma_2 \Rightarrow^* a_1 a_2 \cdots a_{n-1} A_{n-1} \Rightarrow a_1 a_2 \cdots a_n$$

- **対応する動作**: Relating transitions

$$\begin{aligned}(q, a_1 a_2 \cdots a_n, S) &\vdash (q, a_2 \cdots a_n, A_1 \gamma_1)\\ &\vdash (q, a_3 \cdots a_n, A_2 \gamma_2)\\ &\cdots\\ &\vdash (q, a_n, A_{n-1})\\ &\vdash (q, \epsilon, \epsilon)\end{aligned}$$

- **遷移関数**: Transition functions
  - **生成規則** $A \to a\gamma$ **があり、かつその限り**: If and only if there is a production rule $A \to a\gamma$

$$(q, \gamma) \in \delta(q, a, A)$$

例 4.1:

$$G = \langle \{S, A, B\}, \{\mathsf{a}, \mathsf{b}\}, P, S \rangle$$
$$P = \{S \to \mathsf{a}|\mathsf{b}|\mathsf{a}SA|\mathsf{b}SB, A \to \mathsf{a}, B \to \mathsf{b}\}$$

$$S \Rightarrow \mathsf{a}SA \Rightarrow \mathsf{ab}SBA \Rightarrow \mathsf{aba}SABA \Rightarrow \mathsf{abaa}ABA$$
$$\Rightarrow \mathsf{abaaa}BA \Rightarrow \mathsf{abaaab}A \Rightarrow \mathsf{abaaaba}$$

$$M = \langle \{q\}, \{\mathsf{a}, \mathsf{b}\}, \{S, A, B\}, \delta, S, \emptyset \rangle$$

$\delta(q, \mathsf{a}, S) = \{(q, \epsilon), (q, SA)\}$	$S \to \mathsf{a} \mid \mathsf{a}SA$より
$\delta(q, \mathsf{b}, S) = \{(q, \epsilon), (q, SB)\}$	$S \to \mathsf{b} \mid \mathsf{b}SB$より
$\delta(q, \mathsf{a}, A) = \{(q, \epsilon)\}$	$A \to \mathsf{a}$より
$\delta(q, \mathsf{b}, B) = \{(q, \epsilon)\}$	$S \to \mathsf{b}$より

$$
\begin{aligned}
(q, \mathsf{abaaaba}, S) &\vdash (q, \mathsf{baaaba}, SA) \\
&\vdash (q, \mathsf{aaaba}, SBA) \\
&\vdash (q, \mathsf{aaba}, SABA) \\
&\vdash (q, \mathsf{aba}, ABA) \\
&\vdash (q, \mathsf{ba}, BA) \\
&\vdash (q, \mathsf{a}, A) \\
&\vdash (q, \epsilon, \epsilon)
\end{aligned}
$$

空スタックで受理するNPDAに対応する文脈自由文法

CFL corresponding to NPDA accepting by empty stack

$$
\begin{aligned}
M &= \langle Q, \Sigma, \Gamma, \delta, q_0, Z, \emptyset \rangle \\
G &= \langle N, \Sigma, P, S \rangle \\
&\quad q, q' \in Q, A \in \Gamma \text{に対して} [qAq'] \in N
\end{aligned}
$$

- $\forall q \in Q$ に対して $S \to [q_0 Z q]$ を作る
- $(q_1, B_1 \cdots B_k) \in \delta(q, a, A)$ に対して
  - $\forall q_2, \cdots, q_{k+1}$ に対して

$$[qAq_{k+1}] \to a\,[q_1 B_1 q_2]\,[q_2 B_2 q_3] \cdots [q_k B_k q_{k+1}]$$

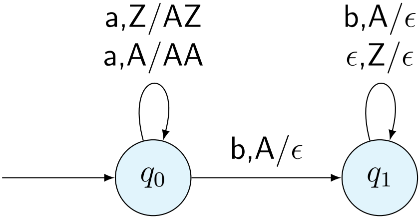
  - ただし $(q_1, \epsilon) \in \delta(q, a, A)$ に対しては

$$[qAq_1] \to a$$

例 5.1:

$$M = \langle \{q_0, q_1\}, \{\mathsf{a}, \mathsf{b}\}, \{A, Z\}, \delta, q_0, Z, \emptyset \rangle$$

$$\begin{aligned}
&\delta(q_0, \mathsf{a}, Z) = \{(q_0, AZ)\}, && \delta(q_0, \mathsf{a}, A) = \{(q_0, AA)\}, \\
&\delta(q_0, \mathsf{b}, A) = \{(q_1, \epsilon)\}, && \\
&\delta(q_1, \mathsf{b}, A) = \{(q_1, \epsilon)\}, && \delta(q_1, \epsilon, Z) = \{(q_1, \epsilon)\}.
\end{aligned}$$



$$
\begin{aligned}
(q_0, \mathsf{aaabbb}, Z) &\vdash (q_0, \mathsf{aabbb}, \mathsf{AZ}) \\
&\vdash (q_0, \mathsf{abbb}, \mathsf{AAZ}) \\
&\vdash (q_0, \mathsf{bbb}, \mathsf{AAAZ}) \\
&\vdash (q_1, \mathsf{bb}, \mathsf{AAZ}) \\
&\vdash (q_1, \mathsf{b}, \mathsf{AZ}) \\
&\vdash (q_1, \epsilon, Z) \\
&\vdash (q_1, \epsilon, \epsilon)
\end{aligned}
$$

対応するCFGを構成: Corresponding CFG

$$G = \langle N, \{\mathsf{a}, \mathsf{b}\}, P, S \rangle$$

- 開始記号: Start symbol

$$S \to [q_0 Z q_0] \mid [q_0 Z q_1]$$

- $\delta(q_0, \mathsf{a}, Z) = \{(q_0, \mathsf{AZ})\}$ より

$$\begin{aligned} [q_0 Z q_0] &\to \mathsf{a}\,[q_0 A q_0]\,[q_0 Z q_0] \mid \mathsf{a}\,[q_0 A q_1]\,[q_1 Z q_0] \\ [q_0 Z q_1] &\to \mathsf{a}\,[q_0 A q_0]\,[q_0 Z q_1] \mid \mathsf{a}\,[q_0 A q_1]\,[q_1 Z q_1] \end{aligned}$$

- $\delta\left(q_0, \mathsf{a}, A\right) = \left\{\left(q_0, \mathsf{AA}\right)\right\}$ より

$$\left[q_0 A q_0\right] \to \mathsf{a}\left[q_0 A q_0\right]\left[q_0 A q_0\right] | \mathsf{a}\left[q_0 A q_1\right]\left[q_1 A q_0\right]$$
$$\left[q_0 A q_1\right] \to \mathsf{a}\left[q_0 A q_0\right]\left[q_0 A q_1\right] | \mathsf{a}\left[q_0 A q_1\right]\left[q_1 A q_1\right]$$

- $\delta(q_0, B, A) = \{(q_1, \epsilon)\}$ より

$$[q_0 A q_1] \rightarrow \mathsf{b}$$

- $\delta(q_1, B, A) = \{(q_1, \epsilon)\}$ より

$$[q_1 A q_1] \rightarrow \mathsf{b}$$

- $\delta(q_1, \epsilon, Z) = \{(q_1, \epsilon)\}$ より

$$[q_1 Z q_1] \rightarrow \epsilon$$

暫定生成規則: Temporary production rules

$$
\begin{aligned}
S &\to [q_0 Z q_0] \mid [q_0 Z q_1] \\
[q_0 Z q_0] &\to \mathsf{a}\,[q_0 A q_0]\,[q_0 Z q_0] \mid \mathsf{a}\,[q_0 A q_1]\,[q_1 Z q_0] \\
[q_0 Z q_1] &\to \mathsf{a}\,[q_0 A q_0]\,[q_0 Z q_1] \mid \mathsf{a}\,[q_0 A q_1]\,[q_1 Z q_1] \\
[q_0 A q_0] &\to \mathsf{a}\,[q_0 A q_0]\,[q_0 A q_0] \mid \mathsf{a}\,[q_0 A q_1]\,[q_1 A q_0] \\
[q_0 A q_1] &\to \mathsf{a}\,[q_0 A q_0]\,[q_0 A q_1] \mid \mathsf{a}\,[q_0 A q_1]\,[q_1 A q_1] \mid \mathsf{b} \\
[q_1 A q_1] &\to \mathsf{b} \\
[q_1 Z q_1] &\to \epsilon
\end{aligned}
$$

生成規則から終端記号に至る非終端記号を探す:

Finding non-terminal symbols leading to terminal symbols

1. **直接終端記号に至る左辺記号**: Non-terminal symbols leading directly to terminal symbols

$$[q_0 A q_1]\,,[q_1 A q_1]\,,[q_1 Z q_1]$$

2. **上記がわかった上で終端記号に至る左辺記号**: Non-terminal symbols leading to terminal symbols based on the above results

$$[q_0 Z q_1]$$

3. **上記がわかった上で終端記号に至る左辺記号**: Non-terminal symbols leading to terminal symbols based on the above results

$$S$$

4. 上記がわかった上で終端記号に至る左辺記号探すが、新たな記号が見つからない: Finding non-terminal symbols leading to terminal symbols based on the above results, but no new symbols are found

終端記号を導く N の要素

- **終端記号を導く N の要素**: Elements of N leading to terminal symbols

$$[q_0 A q_1], [q_1 A q_1], [q_1 Z q_1], [q_0 Z q_1], S$$

- **終端記号を導かない N の要素**: Elements of N not leading to terminal symbols

$$[q_0 Z q_0], [q_1 Z q_0], [q_0 A q_0], [q_1 A q_0]$$

生成規則: Productions

$$
\begin{aligned}
S &\to [q_0 Z q_1] \\
[q_0 Z q_1] &\to \mathsf{a}\,[q_0 A q_1]\,[q_1 Z q_1] \\
[q_0 A q_1] &\to \mathsf{a}\,[q_0 A q_1]\,[q_1 A q_1]\,|\mathsf{b} \\
[q_1 A q_1] &\to \mathsf{b} \\
[q_1 Z q_1] &\to \epsilon
\end{aligned}
$$

導出例: Derivation example

$$
\begin{aligned}
S &\Rightarrow [q_0 Z q_1] \\
&\Rightarrow \mathsf{a}\,[q_0 A q_1]\,[q_1 Z q_1] \\
&\Rightarrow \mathsf{aa}\,[q_0 A q_1]\,[q_1 A q_1]\,[q_1 Z q_1] \\
&\Rightarrow \mathsf{aaa}\,[q_0 A q_1]\,[q_1 A q_1]\,[q_1 A q_1]\,[q_1 Z q_1] \\
&\Rightarrow \mathsf{aaab}\,[q_1 A q_1]\,[q_1 A q_1]\,[q_1 Z q_1] \\
&\Rightarrow \mathsf{aaabb}\,[q_1 A q_1]\,[q_1 Z q_1] \\
&\Rightarrow \mathsf{aaabbb}\,[q_1 Z q_1] \\
&\Rightarrow \mathsf{aaabbb}
\end{aligned}
$$