

5. 文脈自由文法と言語(1): (テキスト5.1)

5.1. 文脈自由文法(CFG; Context Free Grammar)

- 正則言語は言語としては十分な表現能力を持っているとは言えない。

例) $L = \{ 0^n 1^n \mid n \geq 0 \}$... $\{ \varepsilon, 01, 0011, 000111, \dots \}$

$L = \{ \text{括弧の対応が取れている語} \}$...

○ $()$, $(())$, $()()$, $(()())$, ...

× $)$, $)()$, $)()()$, $(()()$, ...

これらは正則言語ではない!!

- 上例の后者は現実の「言語」でも必須の能力
 - 複文(文章の入れ子構造)
 - HTML, LaTeX, C, ...

5. 文脈自由文法と言語(1): (テキスト5.1)

5.1. 文脈自由文法(CFG; Context Free Grammar)

5.1.1. 直感的な例

回文(Palindrome): 前から読んでも後ろから読んでも同じ

例) たけやぶやけた、だんすがすんだ、うついけんしんけいつう

$\Sigma = \{0,1\}$ のとき... $\varepsilon, 0, 1, 00, 11, 000, 010, 101, 111, 0000, \dots$

形式的には... $L_p = \{w \mid w = w^R\}$

L_p は正則言語ではない。

$\Sigma = \{0,1\}$ 上の回文の再帰的定義:

- $\varepsilon, 0, 1$ は回文。
- 回文 w に対して $0w0, 1w1$ は回文。
- この規則で生成できるものだけが回文。

5. 文脈自由文法と言語(1): (テキスト5.1)

5.1. 文脈自由文法(CFG; Context Free Grammar)

5.1.1. 直感的な例

$\Sigma = \{0,1\}$ 上の回文の再帰的定義:

- $\varepsilon, 0, 1$ は回文。
- 回文 w に対して $0w0, 1w1$ は回文。
- この規則で生成できるものだけが回文。

回文を生成する文脈自由文法

1. $P \rightarrow \varepsilon$
2. $P \rightarrow 0$
3. $P \rightarrow 1$
4. $P \rightarrow 0P0$
5. $P \rightarrow 1P1$

5. 文脈自由文法と言語(1): (テキスト5.1)

5.1.2. 文脈自由文法の定義

CFG $G = (V, T, P, S)$

- V : 変数(または非終端記号、文法概念)
 - 書き換えるべき記号
- T : 終端記号
 - 目的とする語を構成するアルファベット
- P : 生成規則
 - 『非終端記号→非終端記号と終端記号の列』という書き換え規則の集まり
- S : 出発記号
 - 最初に出発する非終端記号

$$G_p = \{\{P\}, \{0,1\}, A, P\}$$
$$A: \begin{cases} P \rightarrow \varepsilon \\ P \rightarrow 0 \\ P \rightarrow 1 \\ P \rightarrow 0P0 \\ P \rightarrow 1P1 \end{cases}$$

5. 文脈自由文法と言語(1): (テキスト5.1)

5.1.2. 文脈自由文法の定義

例) $L = \{a, +, (\cdot)\}$ から構成される式

$(a+a), ((a+a)+a+a), (((a)))$, ...

再帰的な定義:

1. a は式
2. E が式なら、 (E) や $E+E$ も式

$G = \{\{P\}, \{a, +, (\cdot)\}, A, P\}$

ただし $A: \begin{cases} P \rightarrow a \\ P \rightarrow (P) \\ P \rightarrow P + P \end{cases}$

$P \rightarrow a \mid (P) \mid P+P$
と書く場合が多い

5. 文脈自由文法と言語(1): (テキスト5.1)

5.1.3. 文法による導出

文法と実際に与えられる語について、、、

- 再帰的推論
文字列(語=終端記号列)から出発記号(非終端記号)
- 導出
出発記号(非終端記号)から文字列(語)



どちらも本質的
には同じ

5. 文脈自由文法と言語(1): (テキスト5.1)

5.1.3. 文法による導出

関係記号[\Rightarrow]

α の中にある非終端記号を一つ、文法 G の生成規則に基づいて書き換えたときに β が得られるとき $\alpha \xRightarrow{G} \beta$ と書く。 G がわかっているときは $\alpha \Rightarrow \beta$ と書く。

関係記号[$\xRightarrow{*}$]

基礎: どんな列に対しても $\alpha \xRightarrow{*}_G \alpha$

再帰: $\alpha \xRightarrow{*}_G \beta$, $\beta \Rightarrow \gamma$ なら、 $\alpha \xRightarrow{*}_G \gamma$ 。

G がわかっているときは省略する。

5. 文脈自由文法と言語(1): (テキスト5.1)

5.1.3. 文法による導出

例)

文法 $G = \{\{P\}, \{a, +, (\cdot)\}, P \rightarrow a \mid (P) \mid P+P, P\}$

に対し、

語 $(a+((a+a)+a))$

の導出は以下の通り:

$P \Rightarrow (P) \Rightarrow (P+P) \Rightarrow (a+P) \Rightarrow (a+(P)) \Rightarrow (a+(P+P))$
 $\Rightarrow (a+((P)+P)) \Rightarrow (a+((P+P)+P)) \Rightarrow (a+((a+P)+P))$
 $\Rightarrow (a+((a+a)+P)) \Rightarrow (a+((a+a)+a))$
 $P^* \Rightarrow (a+((a+a)+a))$

導出の途中で現れる
文字列を文形式と呼ぶ

5. 文脈自由文法と言語(1): (テキスト5.1)

5.1.4. 最左導出と最右導出

非終端記号が複数あった場合に、どの非終端記号から生成規則を適用するか

- 最左導出...もっとも左にある非終端記号から生成規則を適用
- 最右導出...もっとも右にある非終端記号から生成規則を適用

例) $P \Rightarrow (P) \Rightarrow (P+P) \Rightarrow (a+P) \Rightarrow (a+(P)) \Rightarrow (a+(P+P))$
 $\Rightarrow (a+((P)+P)) \Rightarrow (a+((P+P)+P)) \Rightarrow (a+((a+P)+P))$
 $\Rightarrow (a+((a+a)+P)) \Rightarrow (a+((a+a)+a))$

5. 文脈自由文法と言語(1): (テキスト5.1)

5.1.5. ある文法の言語

与えられたCFG $G=(V, T, P, S)$ に対して、 G によって表現される言語 $L(G)$ は

$$L(G) = \{ w \in T^* \mid S \xrightarrow{*}_G w \}$$

と定義できる。

言語 L がある CFG G に対して $L(G)=L$ となるとき、 L は文脈自由言語あるいは CFL (Context Free Language) と呼ばれる。

非終端記号が前後の文脈と関係なく
(=Context Free)書き換えられる

5. 文脈自由文法と言語(1): (テキスト5.1)

5.1.5. ある文法の言語

文法 $G_p = (\{P\}, \{0,1\}, P \rightarrow \varepsilon \mid 0 \mid 1 \mid 0P0 \mid 1P1, P)$ とする。

[定理] $L(G_p)$ は回文の集合である。

[証明] 以下の二つを証明すればよい。

1. $w = w^R$ なら、 $P \xRightarrow{*} w$ であること
2. $P \xRightarrow{*} w$ なら $w = w^R$ であること

5.1.5. ある文法の言語

文法 $G_p = (\{P\}, \{0,1\}, P \rightarrow \varepsilon \mid 0 \mid 1 \mid 0P0 \mid 1P1, P)$ とする。

[定理] $L(G_p)$ は回文の集合である。

[証明] 1. $w = w^R$ なら、 $P \stackrel{*}{\Rightarrow} w$ であること

$|w|$ に関する帰納法による。

[基礎] $|w|=0$ のときは $w = \varepsilon$, $|w|=1$ のときは 0 か 1 であり、いずれも回文である。

[帰納] $|w|=n > 1$ として、 $|w'| < n$ のときは回文 w' は G_p で導出できると仮定。

$w = w^R$ なので、ある文字列 x が存在し、

① $w = 1x1$ か $w = 0x0$ が成立し、かつ

② $x = x^R$ が成立する。

ここで $|x| = |w| - 2 < n$ なので、②と帰納法の仮定より、 $P \stackrel{*}{\Rightarrow} x$ が成立する。

したがって①より

$P \Rightarrow 0P0 \stackrel{*}{\Rightarrow} 0x0 = w$ または $P \Rightarrow 1P1 \stackrel{*}{\Rightarrow} 1x1 = w$ が成立。 12/14

5.1.5. ある文法の言語

文法 $G_p = (\{P\}, \{0,1\}, P \rightarrow \varepsilon \mid 0 \mid 1 \mid 0P0 \mid 1P1, P)$ とする。

[定理] $L(G_p)$ は回文の集合である。

[証明] 2. $P \Rightarrow w$ なら、 $w = w^R$ であること

導出の回数に関する帰納法による。

[基礎] 導出の回数が1回のはきは $w = \varepsilon, 0, 1$ であり、いずれも回文である。

[帰納] w を導出するために規則を n 回 ($n > 1$) 適用したとする。規則を $n - 1$ 回まで適用した場合は回文が生成されると仮定する。

$n > 1$ なので、ある文字列 x が存在し、

① $P \Rightarrow 1P1 \Rightarrow 1x1 = w$ か $P \Rightarrow 0P0 \Rightarrow 0x0 = w$ が成立し、かつ

② x は P から $n - 1$ 回の導出で得られる。

ここで②と帰納法の仮定より、 $x = x^R$ が成立する。

したがって①より $w = w^R$ が成立。

5. 文脈自由文法と言語(1): 演習問題(6)

1. 次の言語を文脈自由文法で表現せよ。(証明不要)

$$L = \{0^n 1^{2n} \mid n > 0\}$$

2. 文法

$$G_b = (\{B\}, \{(\,)\}, B \rightarrow BB \mid (B) \mid \varepsilon, B)$$

はバランスのとれた括弧の列だけを生成することを証明せよ。