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) $L=\{\ 0^n1^n \mid n \geqq 0\}$ …$\{\varepsilon, 01, 0011, 000111, \ldots\}$

$L=\{$括弧の対応が取れている語$\}$ …

○ ( ), (( )), ( )( ), (( )( )),…

×),)(,))( ), (( )((, …

これらは正則言語ではない!!

– 上例の後者は現実の「言語」でも必須の能力

- 複文(文章の入れ子構造)
- HTML, LaTeX, C, …

5. Context Free Grammar (1): (Text 5.1)

5.1. Context Free Grammar (CFG)

- Regular language is not enough to represent some languages

  Ex) $L=\{ 0^n 1^n \mid n \geqq 0\}$ …$\{ \varepsilon ,01,0011,000111,\ldots\}$

  L={Words with balanced parentheses} …

  ○ ( ), (( )), ( )( ), (( )( )),…

  ×),)(,))( ), (( )((, …

  They are not regular.

- The latter example is necessary for real languages.
  - complex sentences (including sentences recursively)
  - 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、おかしがすきすきすがしかお

$\Sigma=\{0,1\}$ のとき… ε, 0, 1, 00, 11, 000, 010, 101, 111, 0000,…

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:

- ε, 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. Context Free Grammar (1): (Text 5.1)

5.1. Context Free Grammar (CFG)

5.1.1. Simple example

Palindrome: word that reads the same backwards as forwards

Ex) Madam I'm adam, rotator, Nurses run…

When Σ ={0,1}… ε , 0, 1, 00, 11, 000, 010, 101, 111, 0000,…

Formally, $L_p=\{w \mid w=w^R\}$

L_p is not a regular language.

A recursive definition of palindromes over Σ ={0,1}:

- ε , 0, and 1 are palindromes.
- For a palindrome w, $0w0$ and $1w1$ are palindromes.
- There are no other palindromes.

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:

- ε, 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. $P \to \varepsilon$
2. $P \to 0$
3. $P \to 1$
4. $P \to 0P0$
5. $P \to 1P1$

5. Context Free Grammar (1): (Text 5.1)

5.1. Context Free Grammar (CFG)

5.1.1. Simple example

A recursive definition of palindromes over $\Sigma=\{0,1\}$:

- ε , 0, and 1 are palindromes.
- For a palindrome w, $0w0$ and $1w1$ are palindromes.
- There are no other palindromes.

A CFG that generates palindromes:

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)

$$G_p = \{\{P\}, \{0,1\}, A, P\}$$

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

5.1.2. 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

5. Context Free Grammar (1): (Text 5.1)

$$G_p = \{\{P\}, \{0,1\}, A, P\}$$

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

5.1.2. Formal Definition of CFG

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

- V: **Variable** (or **Nonterminals, Syntactic category**)
  - The symbols that have to be rewritten.
- T: **Terminals**
  - The alphabets for the language.
- P: **Productions**
  - A set of rewriting rules with a form 'Nonterminal → Sequence of Nonterminals and Terminals'
- S: **Start symbol**
  - The productions start from this nonterminal

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,+,(,)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

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

ただし

$$A:\begin{cases} P \to a \\ P \to (P) \\ P \to P+P \end{cases}$$

$P\to a \mid (P) \mid P+P$
と書く場合が多い

5. Context Free Grammar (1): (Text 5.1)

5.1.2. Formal Definition of CFG

Ex) L={Formulas consisting of a,+,(,)}

$(a+a)$, $((a+a)+a+a)$, $(((a)))$,…

Recursive definition:

1. a is a formula
2. For a formula E, (E) and $E+E$ are formulas.

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

$$A:\begin{cases} P \to a \\ P \to (P) \\ P \to P+P \end{cases}$$

For short, we describe
$P \to a \mid (P) \mid P+P$

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. Context Free Grammar (1):
(Text 5.1)

5.1.3. Derivation by grammar

For a Grammar and given word,

Essentially, they are the same.

- Recursive inference

  From the word (＝sequence of terminals), we reach to the start symbol (nonterminal)

– Derivation

  From the start symbol, we obtain the word (=sequence of terminals)

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[⇒]

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 $\alpha \underset{G}{\Rightarrow} \beta$ と書く。Gがわかっているときは$\alpha \Rightarrow \beta$と書く。

関係記号[$\overset{*}{\Rightarrow}$]

基礎: どんな列に対しても$\alpha \overset{*}{\underset{G}{\Rightarrow}} \alpha$

再帰: $\alpha \overset{*}{\underset{G}{\Rightarrow}} \beta$, $\beta \underset{G}{\Rightarrow} \gamma$なら、$\alpha \overset{*}{\underset{G}{\Rightarrow}} \gamma$。

Gがわかっているときは省略する。

5. Context Free Grammar (1):
(Text 5.1)

5.1.3. Derivation by grammar

Relationship [$\Rightarrow$]

We denote by $\alpha \underset{G}{\Rightarrow} \beta$ when we obtain β from α by replacing a nonterminal symbol in α with a rule in *G*.

If *G* is clear from the context, we simply write $\alpha \Rightarrow \beta$.

Relationship [$\overset{*}{\Rightarrow}$]

Base: For any sequence, $\alpha \overset{*}{\underset{G}{\Rightarrow}} \alpha$

Recursion: If $\alpha \overset{*}{\underset{G}{\Rightarrow}} \beta$ and $\beta \underset{G}{\Rightarrow} \gamma$, $\alpha \overset{*}{\underset{G}{\Rightarrow}} \gamma$.

If *G* is clear, it can be omitted.

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 $G = \{\{P\},\{a,+,(,)\}, P \to a \mid (P) \mid P+P, P\}$

に対し、

語 $(a+((a+a)+a))$

の導出は以下の通り:

導出の途中で現れる文字列を文形式と呼ぶ

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

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

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

$P \overset{*}{\Rightarrow} (a+((a+a)+a))$

5. Context Free Grammar (1): (Text 5.1)

5.1.3. Derivation by grammar

Ex)

For a CFG $G = \{\{P\},\{a,+,(,)\}, P \to a \mid (P) \mid P+P, P\}$,

A derivation for the word $a+((a+a)+a))$ is as follows:

$$P \Rightarrow (P) \Rightarrow (P+P) \Rightarrow (a+P) \Rightarrow (a+(P)) \Rightarrow (a+(P+P))$$
$$\Rightarrow (a+((P)+P)) \Rightarrow (a+((P+P)+P)) \Rightarrow (a+((a+P)+P))$$
$$\Rightarrow (a+((a+a)+P)) \Rightarrow (a+((a+a)+a))$$
$$P \overset{*}{\Rightarrow} (a+((a+a)+a))$$

The string appearing in a derivation is called 'sentential form.'

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) $P \Rightarrow (P) \Rightarrow (\underline{P}+P) \Rightarrow (a+P) \Rightarrow (a+(P)) \Rightarrow (a+(\underline{P}+P))$

$\Rightarrow (a+((\underline{P})+P)) \Rightarrow (a+((\underline{P}+P)+P)) \Rightarrow (a+((a+\underline{P})+P))$

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

5. Context Free Grammar (1): (Text 5.1)

5.1.4. Leftmost and rightmost derivations

When a sentential form consists of two or more nonterminals, how can we apply a production rule?

- Leftmost derivation… we apply a rule to the leftmost nonterminal.
- Rightmost derivation…we apply a rule to the rightmost nonterminal.

Ex) $P \Rightarrow (P) \Rightarrow (\underline{P}+P) \Rightarrow (a+P) \Rightarrow (a+(P)) \Rightarrow (a+(\underline{P}+P))$

$\Rightarrow (a+((\underline{P})+P)) \Rightarrow (a+((\underline{P}+P)+P)) \Rightarrow (a+((a+\underline{P})+P))$

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

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

$$L(G) = \{ w \in T^* \mid S \overset{*}{\underset{G}{\Rightarrow}} w \}$$

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<u>前後の文脈と関係なく</u>
(=Context Free)書き換えられる

5. Context Free Grammar (1): (Text 5.1)

5.1.5. Language defined by grammar

For a given CFG $G=(V, T, P, S)$, the language L(G) represented by G is defined by

$$L(G) = \{ w \in T^* \mid S \overset{*}{\underset{G}{\Rightarrow}} w \}$$

A language L is called Context Free Language (CFL) if there exists a CFG G with $L(G)=L$.

Nonterminals are replaced context freely; that is, a rule for S can be applied not according to the symbols before/after S.

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

5. Context Free Grammar (1): (Text 5.1)

5.1.5. Language defined by grammar

Let G_p be a grammar defined by

G_p=({P}, {0,1}, $P \rightarrow \varepsilon \mid 0 \mid 1 \mid 0P0 \mid 1P1$, P).

[Theorem] $L(G_p)$ is the set of all palindromes.

[Proof] It is sufficient to show the following:

1. $w=w^R$ implies $P \overset{*}{\Rightarrow} w$.

2. $P \overset{*}{\Rightarrow} w$ implies $w=w^R$.

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、$P\overset{*}{\Rightarrow}w$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、$P\overset{*}{\Rightarrow}x$ が成立する。

したがって①より

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

5.1.5. Language defined by grammar

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

[Theorem] $L(G_p)$ is the set of all palindromes.

[Proof] 1. $w=w^R$ implies $P\overset{*}{\Rightarrow}w$:

Induction for $|w|$.

[Base] When $|w|=0$, we have $w=\varepsilon$. When $|w|=1$, we have w=0 or w=1. Thus we have palindromes.

[Inductive step] We suppose that $|w|=n>1$, and any palindrome w' with $|w'|<n$ can be derived by G_p.

Since $w=w^R$, there is a word x such that

① $w=1x1$ or $w=0x0$, and

② $x=x^R$.

Now, since $|x| = |w| - 2 < n$, by the inductive hypothesis with ②, we have $P\overset{*}{\Rightarrow}x$.

Hence by ①, we have

$P\Rightarrow 0P0\overset{*}{\Rightarrow}0x0=w$ or $P\Rightarrow 1P1\overset{*}{\Rightarrow}1x1=w$.

5.1.5. ある文法の言語

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

[定理] $L(G_p)$ は回文の集合である。

[証明] 2. $P \overset{*}{\Rightarrow} w$ 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回のときは $w=\varepsilon$, 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 ー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 $P\Rightarrow 1P1 \overset{*}{\Rightarrow} 1x1=w$ か $P\Rightarrow 0P0 \overset{*}{\Rightarrow} 0x0=w$ が成立し、かつ

② x は P から nー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。

5.1.5. Language defined by grammar

$G_p=(\{P\}, \{0,1\}, P\to \varepsilon \mid 0 \mid 1 \mid 0P0 \mid 1P1, P)$

[Theorem] $L(G_p)$ is the set of all palindromes.

[Proof] 2. $P\overset{*}{\Rightarrow}w$ implies $w=w^R$.

Induction for the number of derivations.

[Base] When the number of derivations is 1, we have $w=\varepsilon$, 0, or 1, and hence they are palindromes.

[Inductive step] We assume that we apply the rules n (n>1) times to derive w. We suppose that we have palindromes when we apply the rules at most $n-1$ times.

Since n>1, there exists a string x such that

① $P\Rightarrow 1P1\overset{*}{\Rightarrow}1x1=w$ or $P\Rightarrow 0P0\overset{*}{\Rightarrow}0x0=w$, and

② x can be derived from P with applying the rules $n-1$ times.

By ② with the inductive hypothesis, we have $x=x^R$.

Hence, by ①, we have $w=w^R$.

お知らせ

- 今日のオフィスアワーは講義

(8) 文脈自由文法と言語(2)

Information

- Today’s Office Hour: Lecture

(8) Context Free Grammar/Language (2)