

形式言語とオートマトン

Note1 言語

2020.4.2作成

2020.4.21update 2021.4.7update

中野眞一 群馬大学

言語（もしくはは形式言語）

言語 = 文字列の**集合**

言語の数学的な**モデル** 3つある。

代数的なモデル

= > **正規表現**

グラフのモデル

= > **オートマトン**

書き換えモデル

= > (文脈自由)**文法**

数学の約束

集合は大文字 要素は小文字 で記述します。

語

文字（もしくは記号）を要素とする集合を **アルファベット** という。

例

$$\Sigma = \{1,2,3,4,5\} = \{x \mid x \text{は} 1 \text{以上} 5 \text{以下の整数}\}$$

$$\Delta = \{a,b,c,d,\dots,z\} = \{x \mid x \text{は任意の英小文字}\}$$

アルファベット Σ の要素を、重複を許して有限個並べた列を Σ 上の語という。有限個なので、3個でも100億個でもよいことに注意!

例) $\Sigma = \{ (,) \}$ 上の語

$()$, $(())$, $()()$, $()()$, $((((($

記述のルール

aa や aaa を a^2 や a^3 と書く。

同様に a を k個並べたものを a^k と書く。

例) $\Sigma = \{ (,) \}$ 上の語 $((((((((= (10$

$(^0 = \varepsilon$ に注意! ε は長さ0の文字列です。(後述。)

語w の記号を逆順に並べて得られる語を wの逆といい、 w^R と書く。

例)

$x = (())$ の逆は $x^R =))(($ である

連接

語 x と語 y をこの順につなげた語を x と y の**連接**(concatenation) といい、 xy と書く。

語 x を k 個連接した語を x^k と書く。

例)

$\Sigma = \{ a, b, c \}$ 上の2つの語 $u = abb$ と $v = bcc$ の連接は $uv = abb\ bcc$ である。

v と u の連接は $vu = bcc\ abb$ である。

(可換で**ない**ことに注意!)

$u^3 = abb\ abb\ abb$

語の長さ

語 w を構成する記号の個数を、この語 w の長さといい、 $|w|$ と書く。

例)

$\Sigma = \{ (,) \}$ 上の語 $x = (())$ の長さは4である。 $|x|=4$ である。

特に長さ0の語を空語という。 ε (イプシロン)で表す。

ε は Σ に含まれないことに注意!

(ちなみに ϕ は要素が0個の集合、空集合)

部分語、接頭語、接尾語

語 w が $w = xyz$ と書けるとき

語 x, y, z は語 w の**部分語**(subword)という。

特に x は w の**接頭語**といい z は w の**接尾語**という。

例) $w=abc$ の部分語は、 $\varepsilon, a, b, c, ab, bc, abc$, の7つである。

(ac は部分語では**ない**ことに注意!)

(ε はすべての語の接頭語、接尾語、部分語になっている!)

Σ^k と Σ^*

Σ^k は Σ 上の長さ k の語の集合を表す。

例)

$\Sigma = \{ (,) \}$ のとき

$$\Sigma^0 = \{ \varepsilon \}$$

$$\Sigma^1 = \{ (,) \}$$

$$\Sigma^2 = \{ ((, ()), (,)) \}$$

Σ^* は Σ 上の語の全体の集合を表す。 Σ の閉包という。

つまり

$$\Sigma^* = \Sigma^0 \cup \Sigma^1 \cup \Sigma^2 \cup \dots$$

言語

Σ 上の言語 (形式言語, formal language) とは、
 Σ^* の部分集合である。

(言語に含まれる語と、含まれない語を厳密に区別すること。)

例) $\Sigma = \{ (,) \}$ のとき

言語 $L = \{ w \mid w \in \Sigma^* \text{ かつ } w \text{ は対応のとれた括弧} \}$

$()()() \in L$ である。

$((())) \in L$ である。

$)()() \in L$ は間違い。

言語の例 (つづき)

例) $\Sigma = \{ A, B, C, \dots, Z, a, b, c, \dots, z, _, 0, 1, 2, \dots, 9 \}$ のとき

言語 $L = \{ p \mid p \in \Sigma^* \text{ かつ } p \text{ は Cプログラムの変数名として許される語} \}$

$\text{sum2} \in L$ である。

$\text{2sum} \in L$ は間違い。(また、予約語にも注意!)

例) $\Sigma = \{ a, b \}$ のとき

言語 $P = \{ a^n b^n \mid n \text{ は } 0 \text{ 以上の整数} \}$

$P = \{ \varepsilon, ab, aabb, aaabbb, aaaabbbb, aaaaabbbbb, \dots \}$

• $= \{ \varepsilon, ab, a^2b^2, a^3b^3, a^4b^4, \dots \}$

言語の演算

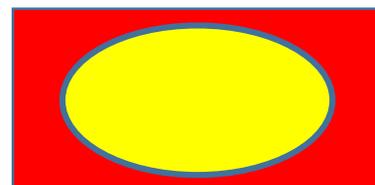
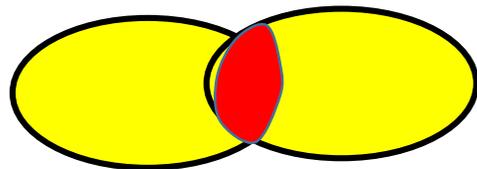
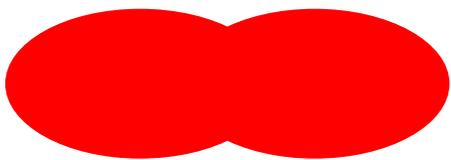
言語から新しい言語を次のようにして作ることができる。

LとMを Σ 上の2つの言語とする。このとき、

LとMの和を $L \cup M = \{x \mid x \in L \text{ または } x \in M\}$ とする。

LとMの共通部分を $L \cap M = \{x \mid x \in L \text{ かつ } x \in M\}$ とする。

Lの補集合を $\sim L = \Sigma^* - L$ とする。



言語の演算 (つづき)

LとMの**連接**(積、product)を $L \cdot M = LM = \{ xy \mid x \in L, y \in M \}$ とする。

L・M と M・L は同じで**ない**かもしれないことに注意しよう。

Lの**n乗**(n-th power)を $L^n = L \cdot L^{n-1}$ とする。

ただし、特にn=0のときは $L^0 = \{ \varepsilon \}$ とする。

特に、 $L^1 = L$ である。

Lの**反転**(鏡像)を $L^R = \{ x^R \mid x \in L \}$ とする。

Lの**閉包**(Kleene closure, スター演算)を $L^* = \bigcup_{n \geq 0} L^n$ とする。

Lの**正閉包**を $L^+ = \bigcup_{n \geq 1} L^n = L^1 \cup L^2 \cup L^3 \cup \dots$

例) $L^4 = L \cdot L \cdot L \cdot L$ である。記号 \cdot は略すこともある。

$L \cdot L \cdot L \cdot L = LLLL$ である。

理解確認クイズ

(1) 語 $a^5b^5\epsilon^5$ の長さはいくつか？

(2) 語 $a^3b^3\epsilon^5$ の接頭語はいくつあるか？

(3) $L = \{ ab, aabb, aaabbb, aaaabbbb, aaaaabbbbb, \dots \}$ を簡潔に記述せよ。

(4) $L = \{ 0, 11, 110, 1001, 1100, 1111, 10010, \dots \}$ を簡潔に記述せよ。

(5) $\Sigma = \{a, b, c\}$ 上の2つの言語 $A = \{a, ab\}$, $B = \{c, bc\}$ とする。
言語 AB の要素を書け。

理解確認クイズ

(6) $\Sigma = \{a,b,c\}$ 上の言語 $A = \{a,ab\}$ とする。言語 A^3 の要素を書け。

(7) $\Sigma = \{a,b,c\}$ 上の言語は何種類あるか？

(8) $\Sigma = \{0,1,2\}$ 上の2つの言語 $L_1 = \{00,01,02\}$, $L_2 = \{\varepsilon\}$ とする。

言語 L_1L_2 の要素を書け。

(9) $\Sigma = \{0,1,2\}$ 上の2つの言語 $L_1 = \{00,01,02\}$, $L_2 = \phi$ とする。

言語 L_1L_2 の要素を書け。(ϕ は空集合)

----(以下上級者向け)----

(10) "対応のとれたカッコ"を簡潔に明確に定義せよ。