〈講　座〉

言語理論の最近の話題 I

笠井琢美* 成島 弘** 野口 広*** 守屋悦朗****

言語理論(linguistics) は広く言語および言語が関連する人間の行動を研究する科学であり，これは自然科学であると同時に人文科学であろう．

しかし，ここで取り上げる言語理論は，狭い意味の言語理論であり，それも主として形態論(morphology)の中で音韻を除き，書かれた文章の文法 (Syntax) を中心としてその数学的理論の概要と最近の研究成果特に日本における研究のわれわれの眼にふれたものにつきのべる．(電子通信学会のオートマトン研究会における成果は貴重なものであるが，特にここではふれないので関連の文献を参照されたい．) 外国における研究についてはたとえば山田尚勇“最近のアメリカのオートマトン理論の紹介”(数理科学, No. 98, August 1971, pp. 66-76) などを参照されたい．ここで扱う範疇の言語理論は非常に形式的であり，形式文法理論とか数理言語理論とかよばれる．

この報告では深い関連をもち，省き難いのではあるが，回路理論，オートマタ理論，情報理論などの成果は一切省いた．また，当然この報告の範囲に入るべき構文解析および Probalistic 言語などについても割愛した．この講座は3回に分けて連載される予定であり，各回に必要な文献を適当な場所に配した．

1. 形式言語

ある有限集合 (以下断りのない場合有限集合は空集合でないとする) Σ をとりあげ，Σ を**字母系**(alphabet)：その要素を**文字** (letter) とよぶ．

Σ の文字の有限列

$x_1x_2\cdots x_n,\ \ x_i\in\Sigma$

を Σ を字母系とする**語** (word) といい，$u,\ v,\ w$ などの文字で示す．

$u=x_1x_2\cdots x_n$

のとき $|u|$ で u に含まれる文字の個数 n を示し(重複を許す) 語 u の**長さ** (length) という．文字を含まない列 ε をも語とみなし，**空語** (the empty word) とよぶ．$|\varepsilon|=0$ である．

空語も含めたすべての語の集合を Σ^* で示す．$u,\ v\in\Sigma^*$ であるとき，

$u=v$ とは $u,\ v$ ともに ε であるか，u も v も ε でないときは，$|u|=|v|\,(\neq 0)$ で $u=x_1\cdots x_n$, $v=y_1\cdots y_n$, $x_i,\ y_i\in\Sigma$, であるとき，$x_i=y_i$ $(i=1,\cdots,n)$ である．

として，語の**相等**を定める．

Σ^* の任意の二つの語 $u,\ v$ に対してその**積**,(concatenatio) uv が次のように定められる．

$u=x_1\cdots x_{|u|},\ \ v=y_1\cdots y_{|v|},\ \ x_i,\ \ y_j\in\Sigma$ であるとき $uv=x_1\cdots x_{|u|}y_1\cdots y_{|v|}$, ただし，$u\varepsilon=\varepsilon u=u$ とする．

この積の演算は言語理論において，最も基本的な演算であり，他の演算と区別したいときは uv の代わりに $u\circ v$ とか $u\cdot v$ とかき，またこのときには**列積**とよぶこともある．

このようにして，集合 Σ^* は $\circ$ を演算として

$u(v\omega)=(uv)w\qquad u,\ v,\ \omega\in\Sigma^*$

$u\varepsilon=\varepsilon u=u$

の条件をみたし (交換法則は一般にみたされない)，Σ を生成元とする自由半群である．$aa=a^2$, $aaabb=a^3b^2$, …などと同じ文字の積は正整数の指数で示される．また任意の $a\in\Sigma$ に対して $a^0=\varepsilon$ と定める．

空集合 ϕ も含めて Σ^* の任意の部分集合 L を Σ を字母系とする**形式言語**という．形式言語の集合は，Σ^* の部分集合族 2^{Σ^*} である．それらの間には次のような集合論的演算が定義される．

Σ を字母系とする形式言語のある族を $\{L_\alpha|\alpha\in\Sigma\}$ とするとき，

和 $\cup_{\alpha\in M}L_\alpha=\{u|u$ は少なくとも一つの L_α の語である$\}$.

共通集合 $\cap_{\alpha\in M}L_\alpha=\{u|u$ はすべての L_α の語である$\}$.

* 京都大学数理解析研究所
** 東海大学理学部数学科
*** 早稲田大学理工学部数学科
**** 電気通信大学電子計算機学科

和，共通集合は $\{L_\alpha\}$ が有限族のとき，それぞれ

$L_1 \cup L_2 \cup \cdots \cup L_n$, $L_1 \cap L_2 \cap \cdots \cap L_n$

で示される．

L_1, $L_2 \in 2^{\Sigma^*}$ に対して次が定まる．

差 $L_1 - L_2 = \{u | u$ は L_1 の語で L_2 の語でない$\}$．

特に，$\Sigma^* - L$ を L の**補集合**という．

以上の演算の結果はもちろん Σ を文字系とする形式言語である．

さらに，形式言語にはつぎのような演算が定義される．

L_1, $L_2 \in 2^{\Sigma^*}$ に対して

積 $L_1L_2 = L_1 \cdot L_2 = \{uv | u \in L_1, v \in L_2\}$．語の積と同じく場合により，$L_1L_2$ は $L_1 \cdot L_2$ で示され，さらに

$L_1(L_2L_3) = (L_1L_2)L_3$

$L\{\varepsilon\} = \{\varepsilon\}L = L$

をみたすが，交換法則は一般にみたされない．空集合 $\phi \in 2^{\Sigma^*}$ については，任意の $L \in 2^{\Sigma^*}$ に対し

$L\phi = \phi L = \phi$

とする．

積を用いて次の演算が定まる．任意の $L \in 2^{\Sigma^*}$ に対して

$L^0 = \{\varepsilon\}$, $L^1 = L$ とし，任意の自然数 $n \geqq 2$ に対して $L^n = L^{n-1} \cdot L$

として L の**累乗**が定まる．特に $L = \phi$ のとき

$L^n = \phi$

とする．次の指数法則がなりたつ．m, n が負でない整数のとき

$(L^m)^n = L^{m \times n}$

$L^m L^n = L^{m+n}$

特に

$$L^* = \bigcup_{n=0}^{\infty} L^n = \{\varepsilon\} \cup L \cup L^2 \cup \cdots \cup L^n \cdots,$$

$$L^+ = \bigcup_{n=1}^{\infty} L^n = L^* - \{\varepsilon\}$$

をそれぞれ L の**星積**あるいはクリーニの演算 (Klene's operation)，＋（プラス）**積**という．

Σ を字母系とする語 $u = x_1x_2 \cdots x_n$ に対して，

$x_nx_{n-1} \cdots x_2x_1$

を u の**逆語** (inverse word, reverse) **あるいは鏡像** (mirror image) といい，u^R で示す．$\varepsilon^R = \varepsilon$ とする．任意の $L \in 2^{\Sigma^*}$ に対し

$L^R = \{u^R | u \in L\}$

を定め，L の**逆**，あるいは**鏡像**という．

$(L^R)^R = L$

$(L_1L_2)^R = L_2^R L_1^R$

である．

Ω を任意の集合とするとき，Ω 上の**（二項）関係** (binary relation) R とは $\Omega \times \Omega$ の部分集合のことである．$(a, b) \in R$ を aRb と書くとき，Ω の元 a_i に対して

a_0Ra_1, a_1Ra_2, $\cdots$, $a_{n-1}Ra_n$

となるとき，これを $a_0R^*a_n$ であると定義し，任意の $a \in \Omega$ に対して，aR^*a とする．R^* を R の**反射推移閉包** (reflexive and transitive closure) という．

2. 生成文法

この理論は 1956 年以来 N. Chomsky によって始められ [5]，S, Ginsburg らによって発展されられたものである．[6][7][8]などはこの分野の標準的な入門書であり，詳しい文献表も示されている．[10] はこの分野（および，これと密接な関連のあるオートマタ理論の分野）の文献を網羅している．[1] はこの分野を概説した論文であり，また初期の重要な論文としては [14]，[15]，[16] がある．その他の入門書，概説書として [2]，[3]，[9] などが参考になろう．最近日本でもこの分野に関する入門書が数多く出ている [11]，[12]，[13]．

Chomsky の基本的アイデアは N. Chomsky "Linguistics, Logics, Psychology, and Computers" Computer Programming and Artificial Intelligence, The University of Michigan, 1958 にのべられている．すなわち，比較的年少の正常な人間の言語を話す能力を次のように考える．

まずとにかく，文章をつぎつぎと生成する器械として，文法が考えられる．

$\boxed{\text{文法}} \longrightarrow$ 文章 (1)

しかし，文章の意味を思うとき，文章の構造が考察されなければならない．そこで，

文法 $\longrightarrow$ $\boxed{\quad}$ $\longrightarrow$ 文章の構造 (2)
文章 $\longrightarrow$

という器械が考えられる．また（1）の出力の文章はある選択をうけて，出力されるものであろうから，特別な文章を生成させる（1）の入力があるに違いない．

この機構は今のところまったくわからないが，一応

$\boxed{?} \longrightarrow$ (1)への入力 (3)

という器械とみられる．

また人間の言語的成長は，文法の確立であり，これは経験によるものであり，次のような器械であるとみることができる．

正しい文章$\longrightarrow$ $\boxed{}$ $\longrightarrow$文法 (4)
誤まった文章$\longrightarrow$

以上のいくつかの言語に関する人間の能力は，とにかく文法を中心としていることがわかる．そこで，Chomsky は (2)〜(4) を背景としつつ，まず (1) の器械をつぎのように形式化した．

$\boxed{\text{句構造文法}} \rightarrow \left\{\text{核文章とその構造}\right\} \rightarrow \boxed{\text{文法的変換}} \rightarrow \left\{\text{文章}\right\}$ (5)

2.1 句構造文法

$G=(N,\Sigma,P,S)$ を**句構造文法** (phrase structure grammar 以下単に文法という) という．ただし，N は有限集合で，その要素は**変数**とよばれ，A, B, C, …などの文字で示される．

Σ は有限集合で，この文法が定める言語の字母系であり，$N\cap\Sigma=\phi$ とする．$N\cup\Sigma$ を V で示す．

P は**プロダクション** (production) とよばれる $x\rightarrow y$ の型の有限集合である．ただし

$x\in N^+$, $y\in V^*$.

S は N の特定の要素．プロダクションは Π: $x\rightarrow y$ で示される．u, v が V^* の元であるとき

$$uxv\underset{G}{\overset{\Pi}{\Rightarrow}}uyv$$

とかく．また，w_0, w_1, …, w_m が V^* の語で

$$w_0\underset{G}{\overset{\Pi_1}{\Rightarrow}}w_1,\ w_1\underset{G}{\overset{\Pi_2}{\Rightarrow}}w_2,\ \cdots,\ w_{m-1}\underset{G}{\overset{\Pi_m}{\Longrightarrow}}w_m \qquad (6)$$

であるとき

$$w_0\underset{G}{\overset{\Pi_1\Pi_2\cdots\Pi_m}{\Longrightarrow}}w_m,\ \Pi_1\Pi_2\cdots\Pi_m\in P^*$$

とかき，列 w_0, w_1, …, w_m を**導出** (derivation) あるいは**生成過程**とよぶ．また便宜のため任意の語 $w\in V^*$ に対して

$$w\underset{G}{\overset{\varepsilon}{\Rightarrow}}w$$

と定める．P^* のある元 α に対し $x\underset{G}{\overset{\alpha}{\Rightarrow}}y$ となることを $x\underset{G}{\overset{*}{\Rightarrow}}y$ で示すこともある．すなわち $\overset{*}{\Rightarrow}$ は関係 $\Rightarrow$ の反射推移閉包である．さらに添字 G は略すこともある．

特に (6) で $w_0=S$ で w_m が Σ^* の語になっていると，w_m はそれ自身以外へ $\Rightarrow$ でうつれない．こうした Σ^* の語 w_m の集合 (Σ を字母系とする形式言語) を G によって生成された**句構造言語** (phrase structure language) といい，$L(G)$ で示す．すなわち

$$L(G)=\{w\in\Sigma^*\,|\,S\underset{G}{\overset{*}{\Rightarrow}}w\}.$$

二つの文法 G_1, G_2 は，$L(G_1)=L(G_2)$ となるとき**弱同値** (weakly equivalent) であるという．

以上に定義した一般の文法を **0 型文法**という．

文法 $G=(N,\Sigma,P,S)$ において，プロダクションが常に

$uAv\rightarrow uyv$

ただし，$A\in N$, $y\in V^+$, u, $v\in V^*$ であるとき G は **1 型文法**あるいは**文脈文法** (context-sensitive grammar) という．この条件はプロダクション

$X\rightarrow Y$, $X\in N^+$, $Y\in V^*$

が

$|X|\leqq|Y|$

の単調性の条件をみたすことと同値であり，1 型文法をこの故に**単調文法**ともいう．

次にプロダクションが常に

$A\rightarrow v$

ただし，$A\in N$, $v\in V^*$

であるとき **2 型文法**あるいは**自由文法** (context-free grammar) という．

さらにプロダクションが

$A\rightarrow aB \qquad A\rightarrow a$

ただし，A, $B\in N$, $a\in\Sigma$ の形をしているとき，G は **3 型文法**，**右線型文法**，あるいは**正則文法** (regnlar grammar) という．

それぞれの文法で定義される言語は，**0 型言語**，**1 型言語**あるいは**文脈言語**，**2 型言語**あるいは**自由言語**，**3 型言語**あるいは**正則言語**または**正則集合**という．

定義から文脈言語と正則集合 $L(G)$ は ε を含まないが，プロダクション $S\rightarrow\varepsilon$ を加え (S はプロダクションの右辺にはあらわれないようにしておく) $L(G)\cup\{\varepsilon\}$ を文脈言語 (正則集合) とみなす場合がある．この場合集合族としては

正則集合族$\subsetneqq$自由言語族$\subsetneqq$文脈言語族$\subsetneqq$0 型言語族

となる．

0 型言語，文脈言語，自由言語，正則集合はそれぞ

れチューリング機械 (Turing machine), 線型有界オートマトン (linear bounded automaton), プッシュダウンオートマトン (pushdown automaton), 有限オートマトン (finite automaton) で受理される集合として特徴づけられる. これらの言語族のカテゴリーと演算の関係はのちにまとめてのべることにする.

また決定問題について, 各集合族について次のような結果がえられている. 不は決定不可能(undecidable), 可は決定可能 (decidable) であることを示す.

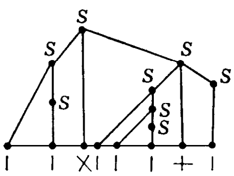
問題	0型	1型 文脈言語	2型 自由言語	3型 正則集合
$L(G)$ は空集合か 有限集合か 無限集合か	不	不	可	可
$L(G)=\Sigma^*$ か	不	不	不	可
$L(G_1)=L(G_2)$ か	不	不	不	可
$L(G_1)\cap L(G_2)$ は空集合か 有限集合か 無限集合か	不	不	不	可
R がある正則集合のとき $L(G)=R$ か	不	不	不	可
$L(G)$ は正則集合か	不	不	不	可

0型言語族は帰納的可算集合族と一致する. 文脈言語は帰納的集合ではあるが, 文脈言語とならない帰納的集合が存在する. 自由言語では生成過程が木で表示され, これを用いて曖昧性が論じられる. たとえば

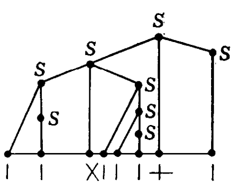
$$G=(S,\Sigma=\{1,\times,+\},P,S)$$

$$P=\begin{Bmatrix} S\to S\times S \\ S\to S+S \\ S\to 1\,S \\ S\to 1 \end{Bmatrix}$$

では, $L(G)\ni 11\times 111+1$ が



(7)



(8)

と, 異なる二つの過程で生成されそれぞれ7, 8を示すので曖昧であるが, ポーランド記法を用いて, 上述の P の代わりに

$$P_P=\begin{cases} S\to\times SS \\ S\to +SS \\ S\to 1\,S \\ S\to 1 \end{cases}$$

を用いれば上記の過程に対応してそれぞれ

$\times 11.+111.1.$　　$+\times 11.111.1.$

という語が生成されて, 曖昧でない言語である.

言語はそれを生成するどのような自由文法をえらんでも, つねにその文法が曖昧であるときに, **曖昧**(ambiguous) であるという. たとえば, $\{a^ib^jc^k \mid i=j$ か $j=k\}$ は曖昧な言語の例である.

2.2 各論

ここでは正則集合, 自由言語の性質をまとめて述べる.

$\boxed{\text{正則集合}}$ はほとんど完璧にその性質が研究しつくされている.

Σ^* における正則集合の族は, 有限集合族を含み, 和, 積, 星積に関して閉じている最小の集合族であり, さらに補集合, 共通部分の演算についても閉じていてブール代数系をなす.

また, 次の3つの命題が同値である.

(i) Uが正則集合である.

(ii) Σ^*に Σ^* を有限個の同値類に分ける同値関係 $\approx$ が定義され, Uはその同値類のいくつかの和集合として示される. ただし, 同値関係 $\approx$ は右 (左) 不変性の性質をみたす. すなわち, U の元 x, y が $x\approx y$ であると, Σ^*の任意の元 z に対して $xz\approx yz$ $(zx\approx zy)$ となる.

(iii) Σ^* の元 x, y に対し $x\equiv y$ であるとは, すべての z, w に対し, $zxw\in U \rightleftharpoons zyw\in U$ となることであると, 関係 $\equiv$ を定めると, $\equiv$ は Σ^* の同値関係で右 (左) 不変であり, その同値類の個数は有限個である.

また次のような公理論がえられている[17], [18].

公理系

Σ 上の**正則形式** (regular expression) を次のように帰納的に定める.

(1) Σ の元および記号 ϕ は正則形式である.

(2) α および β が正則形式ならば $(\alpha+\beta)$, $(\alpha\cdot\beta)$, α^* はそれぞれ正則形式である.

(3) (1)および (2) の有限回の適用によって得られたもののみが正則形式である.

(記号 *, ・, +の順に強いという条件のもので, 括弧は省略できる. また $\alpha\cdot\beta$ の記号・を省略し $\alpha\beta$ と

かく.)

Σ 上の正則形式 α に対して，つぎの規則で Σ^* の部分集合 $\|\alpha\|$ を対応させ $\|\alpha\|$ を正則形式 α によって表示された **Σ-正則集合** (以下単に正則集合) という.

(1) 正則形式 $a \in Z$ に対して$\|a\|=\{a\}$, $\|\phi\|=\phi$ (空集合).

(2) 正則形式 α, β で表示された正則集合をそれぞれ $\|\alpha\|$, $\|\beta\|$ とする. そのとき $\|\alpha\beta\|=\{xy \mid x \in \|\alpha\|,\ y \in \|\beta\|\}$, $\|\alpha+\beta\|=\|\alpha\| \cup \|\beta\|$, $\alpha^*=\{\varepsilon\} \cup \left(\sum_{i=1}^{\infty} \|\alpha^i\|\right)$, (とくに $\|\phi^*\|=\{\varepsilon\}$).

ε-形式とは次のように帰納的に定義される正則形式である.

(1) 任意の正則形式 α に対して，α^* は ε-形式である.

(2) $\alpha+\beta$ は α または β が ε-形式のとき ε-形式である.

(3) $\alpha\beta$ は α, β がともに ε-形式のとき ε-形式である.

正則形式 α, β について，$\|\alpha\|=\|\beta\|$ のとき，それらは**等価**であるといい，$\alpha=\beta$ とかく.

公理

(A_1) $\alpha+(\beta+\gamma)=(\alpha+\beta)+\gamma$

(A_2) $\alpha(\beta\gamma)=(\alpha\beta)\gamma$

(A_3) $\alpha+\beta=\beta+\gamma$

(A_4) $\alpha(\beta+\gamma)=\alpha\beta+\alpha\gamma$

(A_5) $(\alpha+\beta)\gamma=\alpha\gamma+\beta\gamma$

(A_6) $\alpha+\alpha=\alpha$

(A_7) $\phi^*\alpha=\alpha$

(A_8) $\phi\alpha=\alpha$

(A_9) $\alpha+\phi=\alpha$

(A_{10}) $\alpha^*=\phi^*+\alpha^*\alpha$

(A_{11}) $\alpha^*=(\phi^*+\alpha)^*$

(A_{12}) $\alpha^*=\phi^*+\alpha\alpha^*$

推論規則(R I) 正則形式 γ の中にあらわれている正則形式 α をすべて β でおきかえて得られる正則形式を γ' とする. そのとき $\alpha=\beta$ および $\gamma=\delta$ から $\gamma'=\gamma$ および $\gamma'=\delta$ を推論する.

(R II) β が ε-形式でないとき，$\alpha=\alpha\beta+\gamma$ から $\alpha=\gamma\beta^*$ を推論する.

(R III) β が ε-形式でないとき，$\alpha=\beta\alpha+\gamma$ から $\alpha=\beta^*\gamma$ を推論する.

以上の公理と推論規則をもつシステムを公理系 F とよぶ.

定理 公理系 F は無矛盾かつ安全である.

分解定理 *言語 (star event) が研究され次のような結果がえられている [19].

Σ^* の部分集合 L に対し，Σ^* の部分集合 X が存在して $L=X^*$ となるとき L を ***言語**とよび，X を L の**根** (root) とよぶ. Σ^* の部分集合 L と Σ^* の元 u に対して，$\{x \mid ux \in L\}$ を L の u による左商 (left quotient) といい，とくに L が正則集合のとき**微分** (derivative) といい，$\partial_u L$ または $u \backslash L$ であらわす. また，$L=L_1L_2$ ならば $L_1=\{\varepsilon\}$ または $L_2=\{\varepsilon\}$ のとき，L を**素言語** (prime event) という.

定理 (1) 次の (i)〜(iii) は $L(\subseteq \Sigma^*)$ が *言語であるための必要十分条件である.

(i) $L=L^*$, (ii) $L=L^2$, (iii) Σ^* の元 u に対して，$\varepsilon \in \partial_u L \leftrightarrow \partial_u L \supseteq L$

(2) 任意の*言語 L は最小の根 L_I をもち，L_I は一意に決まり，$L_I=(L-\{\varepsilon\})-(L-\{\varepsilon\})^2$ である.

(3) L が正則であるための必要十分条件は L_I が正則であることである.

定理 任意の正則集合 L はある自然数 n が存在して，$L=L_1L_2\cdots L_n$, ここで $L_i (1 \leqq i \leqq n)$ は素言語であるか*言語となる. (この分解は一意でない.)

定義 自由文法 $G=(N, \Sigma, P, S)$が **s-文法**(simple deterministic grammar) であるとはプロダクションがすべて

$$A \rightarrow aB_0 \cdots B_n,\ A,\ B_0,\ \cdots,\ B_n \in N$$
$$a \in \Sigma$$

の形で各 (A, a) に対して上の形のプロダクションが高々一つしかないものをいう. s-文法によって生成される言語を **s-言語** (simple deterministic language) といい，s-言語 L は2つのs-言語の積として表わされないとき**素**であるという.

定理 任意の s-言語 L は一意的に素な s-言語の積 $L=L_1L_2\cdots L_n$ として表わされる[23].

Topics *-自由言語 (star-free event), Noncounting-言語 (event), 群-自由言語などが定義され，Krohn-Rhodes の機械分解定理と関連して，それらの同値性が証明され，またそれらの分類問題が研究されている [20], [21], [22].

なお1次元繰返しオートマトン系において，安定状態 (つぎの時刻でもそこにとどまる状態)，原始状態 (その状態に遷移する状態が悪い状態)，分枝状態 (二つ以上の状態から遷移する状態) などの状態集合が固

定境界条件，環状境界条件などの下で，いずれも正則集合となることが示されている[24]．また有限オートマトンによる自己写像の最大不変集合が正則集合かどうか詳しく調べられている[25]．

言語 L_i, $1\leqq i\leqq n$ が**正則分離可能**であるとは互いに素な正則集合 R_i, $1\leqq i\leqq n$ が存在して $L_i\subseteq R_i$ となることである．この概念は [26] で形式言語の複雑さの理論を展開するために導入されたものであるが，その後 [27] などで形式言語の一般論を研究するためにつかわれている．

自由言語 の研究は 60 年代における言語理論の中心的課題であった．

自由言語は，次のようにして定まるアルゴル型言語と一致する．

Σ を字母系とし，A_1, …, A_n が Σ^* の部分集合の値をとる変数とする．A_i を変数とし，Σ^* の有限集合を係数とし積で結んだ項 $\pi_i(A_1,\cdots,A_n)$ の有限個の和 $f(A_1,\cdots,A_n)=\bigcup_{i=1}\pi_i(A_1,\cdots,A_n)$ を標準関数という．

n 個の標準関数，$f_1(A_1,\cdots,A_n)$, $f_2(A_1,\cdots,A_n)$, …, $f_n(A_1,\cdots,A_n)$ の対

$$f'(A_1,\cdots,A_n)=\{f_1(A_1,\cdots,A_n),\ f_2(A_1,\cdots,A_n),\ \cdots,\ f_n(A_1,\cdots,A_n)\}$$

を ***n* 対標準関数**という．

ここで

$$\alpha^{(0)}=(\alpha_1^{(0)},\cdots,\alpha_n^{(0)})=f(\phi,\cdots,\phi)$$

とおき，$k\geqq 0$ に対して

$$\alpha^{(k+1)}=(\alpha_1^{(k+1)},\cdots,\alpha_n^{(k+1)})=f(\alpha_1^{(k)},\cdots,\alpha_n^{(k)})$$

を定め，各 $i=0,1,\cdots,n$ に対して

$$\alpha_i=\bigcup_{k=0}^{\infty}\alpha_i^{(k)}$$

を定義し，これらを**アルゴル型言語** (ALGOL-like language) とよぶ．

$(\alpha_1,\cdots,\alpha_n)$ は方程式 $(A_1,\cdots,A_n)$ の解である．

任意の自由言語 $L\subseteq\Sigma^*$ は次のように Dyck 言語 D とある正則言語 R と準同型写像によって特徴づけられる．

$\Sigma=\{a_1,\cdots,a_n\}$ とし，$\Sigma'=\Sigma\cup\{a_1',\cdots,a_n'\}$ とする．ここで $P=\{S\to\varepsilon, S\to Sa_iSa_i'S\,|\,1\leqq i\leqq n\}$ をプロダクションとする文法 $G=(\{S\},\Sigma',P,S)$ で生成される自由言語 D を **Dyck 言語**という．すると，

任意の自由言語 $L\subseteq\Sigma^*$ に対して，Σ' を字母系とする Dyck 言語 D とある正則集合 R と，$(\Sigma')^*$ を Σ^* の上へ写す準同型写像 h が存在して

$$L=h(D\cap R)$$

とかける [Chomsky-Schutzenberger の定理]．

また次のような結果がえられている．

[$uvwxy$ 定理] L が自由言語であるとき，次の性質をみたす整数 p, q が定まる．L の $|z|>p$ である語 z はすべて $z=uvwxy$ の形にかける．ただし $|vwx|\leqq q$ で，v と x はともに ε ではない．そしてさらに各 $i\geqq 0$ に対して語 uv^iwx^iy は L の元である．

[Parikh の定理] L が自由言語で $\Sigma=\{a_i|1\leqq i\leqq n\}$ とする．このとき Σ^* より n 次元整数ベクトル空間 $\boldsymbol{N}^n$ への写像 ψ を Σ^* の任意の元 z に対し

$$\psi(z)=(\#_1(z),\ \cdots,\ \#_n(z))$$

で定める．ここで，$\#_i(z)$ は z に現われる文字 a_i の個数を示す．すると

$\psi(L)$ は半線形集合である．

ここで半線形集合とは，線形集合の有限個の和集合のことであり，線形集合とは，ある $\boldsymbol{N}^n$ の点 c と p_1, …, p_r に対して $\left\{x\,|\,x=c+\sum_1^m k_ip_i,\ u_i\in\boldsymbol{N}\right.$ (負でない整数)$\left.\right\}$のことである．

定義 自由文法 $G=(N,\Sigma,P,S)$ が **Chomsky 標準形** (Chomsky normal form) であるとは，プロダクションが次のいずれかの形をしていることである．

$A\longrightarrow BC$; A, B, $C\in N$

$A\longrightarrow a$; $A\in N$, $a\in\Sigma$.

定義 G が **Greibach 標準形**であるとは，プロダクションが全て次の形のときをいう．

$A\longrightarrow aB_1\cdots B_n$, $n\geqq 0$

[標準形定理] 任意の自由言語 L に対して

(1) $L(G)=L-\{\varepsilon\}$ なる Chomsky 標準形文法 G が存在する．

(2) $L(G)=L-\{\varepsilon\}$ なる Greibach 標準形文法 G が存在する．

[自由言語のあるハイアラーキー] [28]

定義 ***pda*** (pushdown automaton) とは $M=(K,\Sigma,\Gamma,\delta,q_0,Z_0,F)$ のことである．ただし

(1) K, Σ, Γ はそれぞれ有限集合で状態集合，入力字母系，pd-字母系である．

(2) $q_0\in K$, $F\subseteq K$, $Z_0\in\Gamma$ でそれぞれ初期状態，終局集合，pd-初期記号である．

(3) δ は $K\times(\Sigma\cup\{\varepsilon\})\times\Gamma$ から $K\times\Gamma^*$ の有限部分集合への写像であり，$K\ni q$, $\Sigma\cup\{\varepsilon\}\ni a$,

$\Gamma-\{Z_0\}\ni A$ に対して $\delta(q,a,A)\subseteq K\times(\Gamma-\{Z_0\})^*$ かつ $\delta(q,a,Z_0)\subseteq K\times Z_0(\Gamma-\{Z_0\})^*$ となる.

$\delta(q,a,A)\ni(q',\beta)$ ならば $(q,aw,\gamma A)\vdash(q',w,\gamma\beta)$, $w\in\Sigma^*$, $\gamma\in\Gamma^*$ とかき, $\vdash^*$ は $\vdash$ の反射推移閉包とする.

pda $M=(K,\Sigma,\Gamma,\delta,q_0,Z_0,F)$ に対して $\mathrm{Null}(M)=\{w\mid\exists q\in F,\ (q_0,w,Z_0)\vdash^*(q,\varepsilon,Z_0)\}$, $T(M)=\{y\mid\exists p\in K, w\in\Sigma^*, (q_0,w,Z_0)\vdash^*(p,\varepsilon,y)\}$ とかく.

定義 $P(M)$ が見かけの次元 n (n は負でない整数) であるとは $A_1,\cdots,A_n\in\Gamma$ が存在して $P(M)\subseteq Z_0A_1^*\cdots A_n^*$ となること. 見かけの二次元のうち最小のものを $P(M)$ の**次元**という.

$(q,w,y)\vdash(q',w',y')$ に対して $y'\neq Z_0$ のときとくに $(q,w,y)\vdash^\Delta(q',w',y')$ とかく. M が **m-turn bounded** であるとは, $P_i\in K$, $y_i\in\Gamma$, $a_i\in\Sigma\cup\{\varepsilon\}$ に対して $(p_i,a_i,y_i)\vdash^\Delta(p_{i+1},\varepsilon,y_{i+1})$, $1\leqq i\leqq r$, $y_1=Z_0$ ならば $\#\{i\mid\exists s,\ |y_i|<|y_{i+1}|,\ 1\leqq t\leqq s, |y_{i+t}|>|y_{i+s+1}|\}\leqq m$, となることである. ここで $\#(A)$は集合 A の要素の個数をあらわす.

定義 $N\cup\{\omega\}\ni n,\ m$ に対して *qda* M が **(n,m)-有界**であるとは, (1) $n=\omega$ であるか, $P(M)$は次元 $n'\leqq m$ であり, かつ (2) $m=\omega$ であるか M は m-turn bounded であることをいう. $D_{m,n}$ で (n,m)-有界な *pda* の族をあらわす.

$N\cup\{\omega\}\ni n,\ m$ に対して $\mathcal{F}_{n,m}=\{L\mid\exists M\in D_{n,m}, L=\mathrm{Null}(M)\}$ とする.

$$\mathcal{F}_{n,\infty}=\bigcup_{m\in N}\mathcal{F}_{n,m},\quad \mathcal{F}_{\infty,m}=\bigcup_{n\in N}\mathcal{F}_{n,m}\ \text{かつ}\ \mathcal{F}_{\infty,\infty}=\bigcup_{m,n\in N}\mathcal{F}_{n,m}\ \text{とする.}$$

$\mathcal{F}_{0,m}=\mathcal{F}_{n,0}$ は正則集合の族であり, $\mathcal{F}_{\omega,\omega}$ は自由言語の族と一致し, $\mathcal{F}_{1,\omega}$ は one-counter 言語の族であり, $\mathcal{F}_{\omega,1}$ は線形言語 (linear language) の * 積となる.

定理 (1) $N\cup\{\omega\}\ni n,m$ に対して $\mathcal{F}_{n,m}$ は full AFL(*) である.

(2) $(N-\{0\})\cup\{\omega,\infty\}\ni n,m$ に対して $n\leqq n'$, $m\leqq m'$ であることが$\mathcal{F}_{n,m}\subseteq\mathcal{F}_{n',m'}$ であるための必要十分条件である.

特に $n<n'$, $m\leqq m'$ か $n\leqq n'$, $m<m'$ なら $\mathcal{F}_{n,m}\subsetneqq\mathcal{F}_{n',m'}$ となる.

[有界自由言語]

定義 Σ 上の言語 L が**有界** (bounded) であるとは Σ^* の元 $w_1, w_2, \cdots, w_n$ が存在して $L\subseteq w_1^*w_2^*\cdots w_n^*$ となることである.

定義 $N^n\supseteq M=\{v_1,v_2,\cdots,v_m\}$ は次の (1) と (2) を満たすとき **stratified** であるという, (1) 各ベクトル v_i は高々 2 つの座標を除いて 0, (2) 0でない座標を 2 つもつ M の任意の二つのベクトル $v_i=(0,\cdots,0,x_k,0,\cdots,0,x_l,0,\cdots,0)$, $v_j=(0,\cdots,0,y_{k'},0,\cdots,0,y_{l'},0,\cdots,0)$ に対しては $k<k'<l<l'$ とも $k'<k<l'<l$ ともならない.

定理 L を有界言語とする. L が自由言語であるための必要十分条件は $\Psi(L)$ が半線形かつstratified であることである.

$N^n\supseteq M$ は, 0でない座標を二つ以上もつ M の任意の二つのベクトルが少なくとも一つの座標において共に 0 ではない要素をもつとき, **pairwise connected** であるという.

$N^n\ni c$, $N^n\supseteq M=\{v_i\mid 1\leqq i\leqq m\}$ に対して, 線形集合$\left\{c+\sum_{i=1}^{m}k_iv_i\mid k_i\in N, 1\leqq i\leqq m\right\}$を $L(c,M)$ であらわす. $\varphi: N^n\to a_1^*a_2^*\cdots a_n^*$ を次のように定められた単射とする, $N^n\ni(k_1,k_2,\cdots,k_n)$ に対して $\varphi(k_1,k_2,\cdots,k_n)=a_1^{k_1}a_2^{k_2}\cdots a_n^{k_n}$, $N^n\supseteq M$ に対して $\varphi(M)=\{\varphi(v)\mid v\in M\}$.

$\mathfrak{S}_n$ を n 次の置換の全体とし, $\tilde{\mathfrak{S}}=\bigcup_{n=1}^{\infty}\mathfrak{S}_n$ とおく. $\tilde{\mathfrak{S}}\ni\alpha$, $\Sigma^*\ni w=w_1w_2\cdots w_n$, $w_i\in\Sigma$, $1\leqq i\leqq n$, に対して $\alpha(w)=w_{\alpha(1)}w_{\alpha(2)}\cdots w_{\alpha(n)}$ と定め, **w の α による置換**とよぶ. $\Sigma^*\supseteq L$ に対して$\{\alpha(w)\mid w\in L,\alpha\in\tilde{\mathfrak{S}}\}$ を $p_\tau(L)$ であらわす. このとき, $p_\tau(L)$ が自由言語となるための条件が, 次のようにえられている [7].

定理 (1) $p_\tau(L)=p_\tau(\varphi(L(c,M)))$ に対してMが1次独立ならば, M が stratified かつ pairwise connected であることが $p_\tau(L)$ が自由言語であるための必要十分条件である.

(2) $\psi(L)=\bigcup_{i=1}^{r}L(c,M_i)$; 各 M_i は stratified かつ pairwise connected ならば $p_\tau(L)$ は自由言語である.

$N^3\supseteq M$ のとき, M は pairwise connected となるので, 次の系をえる. これは Ginsburg [7] によって提出された open problem への答である.

系 $a^*b^*c^*\supseteq L$ に対して, $p_\tau(L)$ が自由言語であることと L が自由言語であることは同値である.

ここで詳しくふれることはしなかったが, 有界自由言語は代数的取り扱いが可能であり, またほとんどの

決定問題は可解である.

[構文構造の複雑さによる(自由)言語の分類]

言語の構文構造の類似性ということを形式的に定義して,言語を類別する研究があり [31] 興味深い.くわしい意味についてはふれないが(原論文を参照されたい)表面的なことを以下簡単に述べておこう.

定義 一般 2 *ndfa* (generalized 2-input non deterministic finted automaton) とは $A=(S, \Sigma_1, \Sigma_2, \delta, s_0, F)$ のことである.ただし

(1) S は空でない有限集合で状態の集合である.

(2) Σ_1, Σ_2 はそれぞれ有限集合で入力字母系,出力字母系である.

(3) δ は (s, M, N, s') 型の要素の有限集合,ここで s, $s'\in S$, M, N はそれぞれ $\Sigma_1{}^*$, $\Sigma_2{}^*$ の正則部分集合,からなる有限集合であり

(4) $s_0\in S$, $F\subseteq S$ はそれぞれ初期記号,受理集合という.

定義 δ の元の有限列 $(s_{j_0}, M_1, M_2, s_{j_1})$, $(s_{j_1}, M_2, N_2, s_{j_2})$, $\cdots$, $(s_{j_{p-1}}, M_p, N_p, s_{j_p})$ を A の**道** (path) とよび, $s_{j_0}\xrightarrow{M_1/N_1}s_{j_1}\xrightarrow{M_2/N_2}s_{j_2}\cdots s_{j_{p-1}}\xrightarrow{M_p/N_p}s_{j_p}$ ($s_{j_0}\overset{M/N}{\Longrightarrow}s_{j_p}$, $M=M_1M_2\cdots M_p$, $N=N_1N_2\cdots N_p$) とかく.この一般 2 $ndfaA$ は次のように**語関係** (word relation) $R(A)$ を定める. $R(A)=\{(u, v)\,|\,s_0\overset{M/N}{\Longrightarrow}s', s'\in F, u\in M, v\in N$ である A の道が存在する$\}$. δ の元が $(s, \{u\}, \{v\}, s')$, $u\in\Sigma_1{}^*$, $v\in\Sigma_2{}^*$, であるとき一般 2 $ndfa$ を単に 2 $ndfa$ とよぶ. 2 $ndfa$ によって定まる語関係を **2 項 transduction** という.語関係が 2 $ndfa$ によって定まることと一般 2 $ndfa$ によって定まることとは同値である.

定義 部分語関数 (partial word function)(語関係が関数の性質, すなわち任意の u に対して $(u, v)\in R(A)$ となる v はたかだか一つ,をみたすとき)である 2 項 transduction を**関数 2 項 transduction** (fbt と略す) という.

すべての fbt のクラスを F_0 とかく. F_0 はつぎの (1), (2) をみたす. (1) f_1, $f_2\in F_0$ ならば $f_1\circ f_2\in F_0$ である (ただし, $\circ$ は関数の合成を示す.) (Elgot and Mejei [30]). (2) 任意の字母系 Σ_1 に対して $\Sigma_1{}^*$ 上の恒等関数も F_0 の元である.

定義 Ω をすべての言語のクラスとする. $f\in F_0$, L_1, $L_2\in\Omega$ に対して f が L_1 の L_2 への**翻訳** (translation) であるとは, (1) $L_1\subseteq f$ の定義域, (2) $f(L_1)\subseteq L_2$, $f(\Sigma^*-L_1)\subseteq\Sigma^*-L_2$ (ただし Σ は任意の字母系を含む可算無限集合である), をみたすことをいい, $L_1\leq_{\mathrm{fbt}}L_2(f)$ とかく. L_1, $L_2\in\Omega$ に対して, $L_1\leq_{\mathrm{fbt}}L_2(f)$ となる F_0 の元 f が存在するとき L_1 は L_2 に **fbt-翻訳可能**といい, $L_1\leq_{\mathrm{fbt}}L_2$ とかく.

F_0 に対する (1), (2) の性質から関係 $\leq_{\mathrm{fbt}}$ は反射的かつ推移的であるが反対称的でない.

定義 L_1, $L_2\in\Omega$ に対して $L_1\leq_{\mathrm{fbt}}L_2$ かつ $L_2\leq_{\mathrm{fbt}}L_1$ ならば L_1 と L_2 とは **fbt-同値**といい $L_1\equiv_{\mathrm{fbt}}L_2$ とかく. L を含む fbt-同値類を $[L]_{\mathrm{fbt}}$ であらわし,すべての同値類の集合を $\Omega/\equiv_{\mathrm{fbt}}$ とかく. $\Omega/\equiv_{\mathrm{fbt}}$ の上に $[L_1]\leq_{\mathrm{fbt}}[L_2]\leftrightarrow L_1\leq_{\mathrm{fbt}}L_2$ によって fbt-翻訳可能関係が定まり,この関係は半順序関係となる.

定理 (1) R が 2 項 transduction ならばつぎの (i), (ii), (iii) をみたす fbt f が存在する. (i) $f\subseteq R$, (ii) f の定義域 $=R$ の定義域, (iii) f は f の定義域のおのおのの元 u に対して $s_0\overset{\{u\}/\{v\}}{\Longrightarrow}s'$, $s'\in F$ なる A の道がただ一つ存在するような 2$ndfa$ $A=(S, \Sigma_1, \Sigma_2, \delta, s_0, F)$ によって定まる.

(2) 可算無限集合のすべての有限部分集合族 (半順序として集合の包含関係をもつ) と順序同型である超線形自由言語の fbt-同値類の集合 (半順序として $\leq_{\mathrm{fbt}}$ をもつ) が存在する.

(3) $\mathcal{L}$ をある言語の族とする.言語 L_0 は, $\mathcal{L}$ に含まれる任意の言語 L に対して $L\leq_{\mathrm{fbt}}L_0$ であるとき $\mathcal{L}$ に対して fbt-universal であるという.すると, 0 型および 1 型の言語の族はそれぞれ fbt-universal な言語を含んでいる.

[W-構造]

[32] において文字列の集合の字の議論を木および W-構造とよばれる木の列の集合に拡張し, 2 次元的対象の"一般正則集合"が研究され,そこでの結果が自由言語に適用されいくつかの興味ある結果がえられている."一般正則集合"のあるクラスは自由言語の構造記述集合のクラスとなり,自由言語のクラスと pda のクラスとの間の"明確なる対応"が与えられ,もう一つの結果は自由文法を二つの簡単な段階に分解することで,はじめに自由文法の書き換え規則を制限することによって"nested set"とよばれる自由言語の部分クラスをえる.つぎに準同型写像を適用することによって,自由言語のクラスをえることができる.このことは Chomsky-Schutzenberger の定理の"refine-

ment" ともなる．なお木-オートマトンとの関連は後に述べられる予定である．

[1次元繰返しオートマトン系の言語受理能力]

[33]において，並列型繰返し系，直列型繰返し系の言語受理能力が考察され，次の結果がえられている．

(1) 確定自由言語，線形自由言語はともに狭義に安定な，1次元1方向並列型繰返し系で受理されること，(2) 1次元直列型繰返し系で受理できない，線形でかつ確定の自由言語が存在すること，(3) 自由言語は広義に安定な1次元1方向並列繰返し系で受理されること，(4) 確定線形有界オートマトン (linear bounded automaton) で受理される言語のクラスは，安定な1次元2方向並列繰り返し系で受理されるクラスに一致すること，などである．さらに [34]，[35] では1次元繰返しオートマトン系による言語の実時間認識の問題が研究されている．

参考文献 I

1) A. V. Aho and J. D. Ullman, The Theory of languages, Math. Systems Theory 2 (1968), 97-125.
2) M.A. Arbib, Theories of Abstract Automata, Prentice-Hall, 1969.
3) Y. Bar-Hillel, Language and Information, Selected essays on their theory and applications, Addison-Wesley, 1964.
4) B. Brainerd, Introduction to the Mathematics of Language Study, Elsever, 1971.
5) N. Chomsky, Three models for the description of languages, IRE Trans. Inform. Theory IT-2 (1956), 113-124.
6) M. Gross and A. Lentin, Introduction to Formal Grammars, Springer, 1970 (相沢他訳「数理言語入門」，東京図書，1972).
7) S. Ginsburg, The Mathematical Theory of Context-Free Languages, McGraw-Hill, 1966.
8) J.E. Hopcroft and J.D. Ullman, Formal Languages and Their Relation to Automata, Addison-Wesley, 1969 (野崎・木村他訳「言語理論とオートマトン」，サイエンス社，1971).
9) R. Kurki-Suonio, Computability and Formal Languages, Auerbach, 1971.
10) D. Wood, Bibliography 23; Formal Language Theory and Automata Theory, Computing Review 11 (1970).
11) 本多波雄，オートマトン・言語理論，コロナ社，1972.
12) 栗原俊彦・吉田 将，言語理論，共立出版，1970.
13) 野口 宏，数理言語入門，ダイヤモンド社，1970.
14) Y. Bar-Hillel, M. Perles and E. Shamir, On formal properties of simple phrase structure grammars, Z. Phonetik Sprachwiss Kommunikat 14 (1961), 143-172.
15) N. Chomsky, On certain formal properties of grammars, Information and Control 2 (1959), 137-167.
16) N. Chomsky, Formal Properties of Grammars, in D. Luce, R. Bush and E. Galenter (eds.), " Handbook of Mathematical Psychology ", John Wiley and Sons, Inc., 1963.
17) A. Salomaa, Two complete axiom systems for the algebra of regular events, J. Assoc. Comput. Mach. 13 (1966), 158-169.
18) 高須 達，論理設計概論—電子計算機の理解のために—，裳華房，1967.
19) J.A. Brozowski and R. Cohen, On accompositions of regular events, J. Assoc. Comput. Mach. 16 (1969), 132-144.
20) A.R. Mayer, A note on star-free events, J. Assoc. Comput. Mach. 16 (1969), 220-225.
21) R.S. Cohen and J.A. Borozowski, Dot-depth of star-free events, J. Comput. System Sci. 5 (1971), 1-16.
22) J. A. Borozowski, K. Čulík II and A. Gadrielian, Classification of noncounting events, J. Comput. System Sci. 5 (1971), 41-53.
23) 房岡 璋，Simple Deterministic Language の素元積への分解について，信学会論文誌，54-C (1971), 132-139.
24) Y. Kobuchi and H. Nishio, Some regular state sets in the system of one-dimensional iterative automata, J. Information Sci. 5 (1973), 199-216.
25) 高橋英之，Finite Automaton による自己同型写像の最大不変集合について，数理解析研講究録，123 (1971), 32-42.
26) K. Lobayashi, Structural complexity of contexf-free languages, Information and Control 18 (1971), 299-310.
27) S. Arisaka, On regular separation of languages, Kyushu Univ. Research Institute of Fundamental Information Science Research Report No. 26 (1972).
28) S. A. Greibach, An infinite hierarchy of context free languages, J. Assoc. Comput. Mach. 16 (1969).
29) T. Ohshiba, On permuting letters of words in context-free languages, Information and Control 20 (1972), 405-409.
30) C. C. Elgot and J. E. Mezei, On relations difined by genralized finite automata, IBM J. Res. Develop. 9 (1965), 47-68.

31) K. Kobayashi, Classification of formal languages by functional binary transduction, Information and Control 15 (1969), 95–109.

32) M. Takahashi, Regular sets of strings, trees and W-structures, Ph. D. dissertation, Pensylvania Univ., 1972.

33) 嵩 忠雄・藤井 護, 一次元繰返論理回路系の能力について, 信学会論文誌 51C (1968), 167–176.

34) S. N. Cole, Real-time computation by n-dimensional iterative arrays of finite state machines, IEEE Trans. Computers C–18 (1969), 349–365.

35) A.R. Smith III, Real-time language recognition by one-dimensional cellular automata, J. Comput. System Sci 6 (1972), 233–253.

(昭和 48 年 3 月 1 日受付)