非終端記号を導入した基本形式体系について

小出智彦[*]　　篠原歩[†]

2009 年 7 月 22 日 (水)

1　はじめに

基本形式体系 (Elementary Formal System, EFS) は Smullyan [5] によって導入された文字列を対象とする一種の論理プログラムである．その後有川ら [1, 3] は EFS を用いて形式言語理論の研究を行い，チョムスキー階層と対応する EFS の部分クラスなどを示した．しかし，言語を表現する文法や機械と対応する EFS を容易に表現できるかというとそうとは限らない．例えば，文脈依存文法と長さ限定 EFS はともに文脈依存言語を生成するが，与えられた文脈依存文法と等価な長さ限定 EFS を見つけるのは容易ではない．また，チョムスキー階層以外の言語クラスの文法や機械を表現するのも困難である．

本研究では EFS に新たに非終端記号を導入した非終端記号付き基本形式体系 (Elementary Formal System with Nonterminals, NEFS) を定義する．これに文法や機械に対応する NEFS をより簡単に表現できることを示す．また，NEFS に EFS の場合と同じ制限を加えた場合の表現力について論じる．

2　非終端記号付き基本形式体系の定義

2.1　基本形式体系言語

Σ, X, Π を互いに交わらない集合とする．Σ と Π は有限集合である．Σ の要素は終端記号と呼ぶ．Π の要素を**述語記号**と呼ぶ．それぞれの述語記号には**引数**と呼ばれる自然数が対応付けられている．引数が n である述語記号 p を p/n で表す．X は可算無限集合であり，X の要素を**変数**と呼ぶ．

$(\Sigma \cup X)^*$ の要素を**項**あるいは**パターン**と呼ぶ．Σ^* からなる項を**基礎項**と呼ぶ．Π 中の述語記号 p/n と n 個の項 $\tau_1, ..., \tau_n$ に対し，$p(\tau_1, ..., \tau_n)$ を**原子論理式**，または単に**アトム**と呼び，n 個の項が全て基礎項の場合には**基礎アトム**と呼ぶ．

$A_1, ..., A_i, B_1, ..., B_j$ をアトムとするとき，次のような“ $\leftarrow$ ”を挟んだアトムの並びを**節**と呼ぶ．

$$A_1, ..., A_i \leftarrow B_1, ..., B_j \qquad (i, j \geq 0)$$

“ $\leftarrow$ ”の左側，すなわち $A_1, ..., A_i$ の部分を節の**頭部**，右側の $B_1, ..., B_j$ の部分を本体と呼ぶ．$i = 1$ となる節を**確定節**と呼ぶ．

基本形式体系は 3 項組 (Σ, Π, Γ) によって定義される．Γ は Σ と Π から構成可能な確定節の有限集合で**公理**と呼ばれる．

互いに異なる変数 x_k $(1 \leq k \leq n)$ と任意のパターン π_k に対し，$\{x_1 := \pi_1, ..., x_n := \pi_n\}$ の形をした有限集合を**代入**と呼ぶ．任意のパターン τ と任意の代入 $\theta = \{x_1 := \pi_1, ..., x_n := \pi_n\}$ に対して，τ 中に現れる全ての変数 x_i をそれぞれ対応する π_i によって同時に置き換えて得られるパターンを $\tau\theta$ と書く．アトム $p(\pi_1, ..., \pi_n)$ に対しては

$$p(\pi_1, ..., \pi_n)\theta = p(\pi_1\theta, ..., \pi_n\theta)$$

と，節 $A \leftarrow B_1, ..., B_j$ に対しては

$$(A \leftarrow B_1, ..., B_j)\theta = A\theta \leftarrow B_1\theta, ..., B_j\theta$$

[*]東北大学大学院情報科学研究科
[†]第 1 著者に同じ

と書く．項（アトム，節）E と代入 θ に対して $E\theta$ を E の**代入例**と呼ぶ．$E\theta$ が終端記号のみからなる場合には**基礎代入例**と呼ぶ．

$T = (\Sigma, \Pi, \Gamma)$ が EFS であるとする．Γ と節 C に対して関係 $\Gamma \vdash C$ を次のように定義する．

1. $C \in \Gamma$ ならば $\Gamma \vdash C$.
2. $\Gamma \vdash C$ ならば任意の代入 θ に対して $\Gamma \vdash C\theta$.
3. $\Gamma \vdash A \leftarrow B_1, ..., B_j, B_{j+1}$ かつ $\Gamma \vdash B_{j+1} \leftarrow$ ならば $\Gamma \vdash A \leftarrow B_1, ..., B_j$.

$\Gamma \vdash C$ ならば C は Γ から**証明可能**であるという．EFS T と述語記号 $p/1$ に対して $L(T, p) = \{w \in \Sigma^* \mid \Gamma \vdash p(w) \leftarrow\}$ と定義する．$L = L(T, p)$ なる EFS T と述語 $p/1$ が存在するとき，言語 L を**基本形式体系言語**と定義する．

例 1. *EFS* $T = (\{a, b, c\}, \{p/1, q/3\}, \Gamma)$ を考える．

$$\Gamma = \left\{ \begin{array}{c} p(xyz) \leftarrow q(x, y, z) \\ q(ax, by, cz) \leftarrow q(x, y, z) \\ q(a, b, c) \leftarrow \end{array} \right\}$$

EFS T と述語記号 p によって定義される言語は $L(T, p) = \{a^n b^n c^n \mid n \geq 1\}$.

2.2 非終端記号付き基本形式体系言語

続いて非終端記号付き基本形式体系の定義を行う．N は非終端記号の有限集合で Σ, Π, X とは交わらないものとする．非終端記号は長さを 0 と定義する．非終端記号を導入する前の定義を拡張し，$(N \cup \Sigma \cup X)^*$ の要素を項とする．$(N \cup \Sigma)^*$ からなる項を**非終端記号付き基礎項**と呼ぶ．項が全て非終端記号付き基礎項であるようなアトムを**非終端記号付き基礎アトム**と呼ぶ．

非終端記号付き基本形式体系は 4 項組 (N, Σ, Π, Γ) によって定義される．Γ は N, Σ, Π から構成可能な確定節の有限集合である．

$T = (N, \Sigma, \Pi, \Gamma)$ が NEFS であるとする．Γ と節 C に対して関係 $\Gamma \vdash_N C$ を次のように定義する．

1. $C \in \Gamma$ ならば $\Gamma \vdash_N C$.
2. $\Gamma \vdash_N C$ ならば任意の代入 θ に対して $\Gamma \vdash_N C\theta$.
3. $\Gamma \vdash_N A \leftarrow B_1, ..., B_j, B_{j+1}$ かつ $\Gamma \vdash_N B_{j+1} \leftarrow$ ならば $\Gamma \vdash_N A \leftarrow B_1, ..., B_j$.

$\Gamma \vdash_N C$ ならば C は Γ から**証明可能**であるという．NEFS T と述語記号 $p/1$ に対して $L(T, p) = \{w \in \Sigma^* \mid \Gamma \vdash_N p(w) \leftarrow\}$ と定義する．$L = L(T, p)$ なる NEFS T と述語 $p/1$ が存在するとき，言語 L を**非終端記号付き基本形式体系言語**と定義する．

NEFS は EFS を拡張したものであるため例 1 のような EFS も NEFS で書くことが可能である．加えて，NEFS では次のような書き方も可能である．

例 2. *NEFS* $T = (\{S, B\}, \{a, b, c\}, \{p/1\}, \Gamma)$ を考える．

$$\Gamma = \left\{ \begin{array}{c} p(xabcy) \leftarrow p(xSy) \\ p(xaSBcy) \leftarrow p(xSy) \\ p(xBcy) \leftarrow p(xcBy) \\ p(xbby) \leftarrow p(xbBy) \\ p(S) \leftarrow \end{array} \right\}$$

NEFS T と述語記号 p によって定義される言語は $L(T, p) = \{a^n b^n c^n \mid n \geq 1\}$.

また次の例は $\{a^n b^n c^n \mid n \geq 1\}$ を生成する単調文法 [2] である．

例 3. 単調文法 $G = (\{S, B\}, \{a, b, c\}, P, S)$ を考える．

$$P = \left\{ \begin{array}{c} S \rightarrow abc \\ S \rightarrow aSBc \\ cB \rightarrow Bc \\ bB \rightarrow bb \end{array} \right\}$$

文法 G によって定義される言語は $L(G) = \{a^n b^n c^n \mid n \geq 1\}$.

例 2 の公理と例 3 の生成規則に注目してみると，これらがきれいに対応づいていることがわかる．一方，通常の EFS では単調文法と対応がつくような公理を記述することはできない．

2.3 EFS と NEFS の部分クラス

EFS や NEFS は言語を定義するには一般的すぎる．そこでこれらにいくつかの条件を加えた部分クラスを導入する．

アトム A に対して，A 中に現れる全ての変数の集合を $v(A)$ で表す．パターン π に対して，π の文字列としての長さを $|\pi|$ で表し，π における変数 x の出現回数を $o(x,\pi)$ で表す．さらに，アトム $p(\pi_1,\pi_2,...,\pi_n)$ に対して，

$$|p(\pi_1,\pi_2,...,\pi_n)| = |\pi_1| + ... + |\pi_n|$$
$$o(x,p(\pi_1,\pi_2,...,\pi_n)) = o(x,\pi_1) + ... + o(x,\pi_1)$$

と定義する．

- 節 $A \leftarrow B_1,B_2,...,B_n$ が**変数限定**であるとは，$i=1,...,n$ に対して $v(A) \supseteq v(B_i)$ が成り立つことをいう．つまり，変数限定節では本体に出現する変数は全て頭部に出現している．変数限定節だけからなる公理をもつ EFS を**変数限定 EFS**，NEFS を**変数限定 NEFS** と呼ぶ．
- 節 $A \leftarrow B_1,B_2,...,B_n$ が**長さ限定**であるとは任意の代入 θ に対して

  $$|A\theta| \geq |B_1\theta| + |B_2\theta| + ... + |B_n\theta|$$

  が成り立つことをいう．長さ限定節だけからなる公理をもつ EFS を**長さ限定 EFS**，NEFS を**長さ限定 NEFS** と呼ぶ．
- 公理中の全ての節が次の形をしている長さ限定 EFS を**単純 EFS**，NEFS を**単純 NEFS** という．

  $$q(\pi) \leftarrow q_1(x_1),...,q_n(x_n) \qquad (n \geq 0)$$

  ただし，π は任意のパターンであり，$x_1,...,x_n$ は π に出現する互いに異なる変数である．定義から明らかなように，述語記号はすべて 1 引数である．
- 項 π が正規であるとは，任意の変数 x に対して $o(x,\pi)$ が成り立つことである．公理中の全ての節の頭部のパターンが正則であるような単純 EFS を**正規 EFS**，NEFS を**正規 NEFS** と呼ぶ．
- 公理中の任意の節が次の形をしている EFS を**片側線形 EFS**，NEFS を**片側線形 NEFS** と呼ぶ．

  $$p(\pi) \leftarrow \quad (\pi \text{は正規パターン})$$
  $$p(ux) \leftarrow q(x) \quad (u \in \Sigma^+, x \in X)$$

表 1: EFS と NEFS の各部分クラスが定義する言語のクラス

	EFS	NEFS
変数限定	$\mathcal{VB}$-$\mathcal{EFS}$	$\mathcal{VB}$-$\mathcal{NEFS}$
長さ限定	$\mathcal{LB}$-$\mathcal{EFS}$	$\mathcal{LB}$-$\mathcal{NEFS}$
正規	$\mathcal{REG}$-$\mathcal{EFS}$	$\mathcal{REG}$-$\mathcal{NEFS}$
片側線形	$\mathcal{OSL}$-$\mathcal{EFS}$	$\mathcal{OSL}$-$\mathcal{NEFS}$

例 1 の EFS は長さ限定 EFS，例 2 の NEFS は長さ限定 NEFS である．それぞれの EFS や NEFS が定義する言語のクラスを表 1 のように書くことにする．Arikawa ら [3] は $\mathcal{VB}$-$\mathcal{EFS}$,$\mathcal{LB}$-$\mathcal{EFS}$,$\mathcal{REG}$-$\mathcal{EFS}$,$\mathcal{OSL}$-$\mathcal{EFS}$ がそれぞれ帰納的可算言語，文脈依存言語，文脈自由言語，正規言語のクラスと一致することを示した (図 1)．

3 NEFS 言語のクラス

NEFS は EFS に非終端記号を新たに追加したもので，一見すると表現力が高まっているように見える．しかし，実際は必ずしもそうとは限らない．

定理 1. $\mathcal{VB}$-$\mathcal{NEFS}$ は $\mathcal{VB}$-$\mathcal{EFS}$ と等価である．

証明. 任意の変数限定 EFS は $N=\emptyset$ である変数限定 NEFS によって模倣することが可能である．よって $\mathcal{VB}$-$\mathcal{EFS} \subseteq \mathcal{VB}$-$\mathcal{NEFS}$ となる．

次に $\mathcal{VB}$-$\mathcal{EFS} \supseteq \mathcal{VB}$-$\mathcal{NEFS}$ となることを示す．$L=L(T,p)$ となる変数限定 NEFS $T=(N,\Sigma,\Pi,\Gamma)$

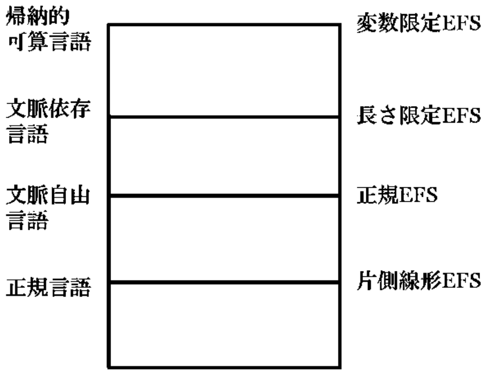

図 1: EFS 言語の部分クラスとチョムスキー階層の各言語クラスの対応

に対して，変数限定 EFS $T' = (\Sigma', \Pi', \Gamma')$ を以下のように構成する．

$$\Sigma' = N \cup \Sigma$$
$$\Pi' = \Pi \cup \{p_c, p_a \mid p_c, p_a \notin \Pi\}$$
$$\Gamma' = \Gamma \cup \{p_c(\epsilon) \leftarrow\} \cup \{p_c(ax) \leftarrow p_c(x) \mid a \in \Sigma\} \cup \{p_a(x) \leftarrow p(x), p_c(x)\}$$

このとき，$L(T,p) = L(T', p_a)$ が成り立つことを証明すればよい．ここで

$$A = \{w \in (N \cup \Sigma)^* \mid \Gamma \vdash_N p(w) \leftarrow\}$$
$$B = \{w \in \Sigma'^* \mid \Gamma \vdash p(w) \leftarrow\}$$

とすると，$A = B$ となることは明らかである．よって $L(T', p_a)$ が $B \cap \Sigma^*$ と等しくなればよい．定義より述語記号 p_c の項 τ は $\tau \in \Sigma^*$ である．節 $p_a(x) \leftarrow p(x), p_c(x)$ により

$$\{w \in \Sigma'^* \mid \Gamma \vdash p_a(w) \leftarrow\} = B \cap \Sigma^*$$

となる．よって $L(T,p) = L(T', p_a)$ となる．

□

系 1. 言語 $L \subseteq \Sigma^*$ が帰納的可算言語であるための必要十分条件は，$L = L(T,p)$ なる変数限定 *NEFS* T が存在することである．

次に長さ限定 NEFS について考える．

補題 1. 節 $A \leftarrow B_1, ..., B_n$ が長さ限定であるための必要十分条件は，

$$|A| \geq |B_1| + ... + |B_n|$$

かつ，任意の変数 x に対して

$$o(x, A) \geq o(x, B_1) + ... + o(x, B_n)$$

が成り立つことである．

これは [3] の Lemma2.1 と同様にして証明可能である．補題 1 の 2 つめの不等式から，全ての長さ限定 NEFS は変数限定 NEFS であることがわかる．

定理 2. 言語 $L \subseteq \Sigma^*$ が帰納的可算言語であるための必要十分条件は，$L = L(T,p)$ なる長さ限定 *NEFS* T が存在することである．

証明. 補題 1 より $\mathcal{LB}$-$\mathcal{NEFS} \subseteq \mathcal{VB}$-$\mathcal{NEFS}$ が成り立つ．系 1 より $\mathcal{LB}$-$\mathcal{NEFS}$ は帰納的可算言語のクラスを超えない．

次に任意の帰納的可算言語 L が長さ限定 NEFS によって定義できることを示す．ここでは帰納的可算言語のクラスが句構造文法で生成できる言語のクラスと等しい [4] ことを利用する．$L = L(G)$ なる句構造文法 $G = (N, \Sigma, P, S)$ を考える．N は非終端記号の集合，Σ は終端記号の集合，P は生成規則の集合，$S \in N$ は開始記号である．P は n 個の生成規則からなり，$P = \{\alpha_i \to \beta_i \mid 1 \leq i \leq n\}$ とする．このとき，長さ限定 NEFS $T = (N', \Sigma', \Pi, \Gamma)$ を以下のように構成する．

$$N' = N \cup \{N_a \mid a \in \Sigma\}$$
$$\Pi' = \Pi \cup \{p_a\}$$

ここで関数 $f : (\{\epsilon\} \cup N \cup \Sigma) \to \{\epsilon\} \cup N'$ を次のように定義する．

$$f(a) = \begin{cases} N_a & a \in \Sigma\text{のとき} \\ a & \text{それ以外のとき} \end{cases}$$

文字列 $a_1...a_m \in (N \cup \Sigma)^+$ に対して $f(a_1...a_m) = f(a_1)...f(a_m)$ とする．また f の逆関数は次のようになる．

$$f^{-1}(B) = \begin{cases} c & B(=N_c) \notin N \text{ のとき} \\ B & \text{それ以外のとき} \end{cases}$$

Γ を次のように構成する．

$$\begin{aligned}\Gamma' =& \{p(xf(\beta_i)y) \leftarrow p(xf(\alpha_i)y) \mid 1 \le i \le n\} \\ & \cup \{p(S) \leftarrow\} \\ & \cup \{p_a(x) \leftarrow p(x)\} \\ & \cup \{p_a(xay) \leftarrow p_a(xN_ay) \\ & \quad \mid a \in \Sigma, N_a \in N' \setminus N\}\end{aligned}$$

Γ' 中の要素は確かに全て長さ限定の条件を満たしている．このとき，$L(G) = L(T, p_a)$ となることを証明すればよい．

P 中の生成規則による導出は節

$$\{p(xf(\beta_i)y) \leftarrow p(xf(\alpha_i)y) \mid 1 \le i \le n\}$$

によって模倣が可能である．よって，

$$A = \{w \mid f^{-1}(w) \text{ は } G \text{ の文形式 }\}$$
$$B = \{w' \mid \Gamma' \vdash_N p(w') \leftarrow\}$$

とすると $A = B$ となる．B の要素のうち，N の要素を含む文字列は N の要素を除去できないので $L(T, p_a)$ には含まれない．N の要素を含まない文字列は，文字列中の全ての非終端記号を節

$$\{p_a(xay) \leftarrow p_a(xN_ay) \mid a \in \Sigma, N_a \in N' \setminus N\}$$

によって終端記号に置き換えられる．定義より全ての文字が終端記号になった文字列だけ $L(T, p_a)$ に含まれる．すなわち，

$$L(T, p_a) = \{w \in \Sigma \mid f(w) \in B\}$$

となる．一方 $L(G)$ は次のようになる．

$$L(G) = \{w \in \Sigma \mid f(w) \in A\}$$

よって $L(G) = L(T, p)$ が成り立つ． □

表 2: チョムスキー階層の各言語クラスと NEFS 言語の各クラスの対応

	EFS	NEFS
帰納的可算	変数限定	変数限定 長さ限定
文脈依存	長さ限定	
文脈自由	正規	正規（予想）
正規	片側線形	片側線形（予想）

正規 NEFS や片側線形 NEFS のクラスについては，節の形の制限が厳しいため正規 EFS や片側線形 EFS と比べても表現力が変わらないと予想される．

予想 1. *1.* $\mathcal{REG}$-$\mathcal{NEFS}$ は $\mathcal{REG}$-$\mathcal{EFS}$ と等価である．

2. $\mathcal{OSL}$-$\mathcal{NEFS}$ は $\mathcal{OSL}$-$\mathcal{EFS}$ と等価である．

4 まとめと今後の課題

本研究では非終端記号付きの基本形式体系を導入し，変数限定 NEFS，長さ限定 NEFS によって定義される言語のクラスが帰納的可算言語と等価であることを示した．また他の NEFS によって定義される言語のクラスについては予想を提示した．(表 2)

今後の課題はチョムスキー階層以外の文法や機械について，従来とは異なる条件付けによってそれらと対応する NEFS を見つけることである．また，同じクラスの言語を生成する文法や機械を NEFS を用いて対応付けることである．

参考文献

[1] Setsuo ARIKAWA. Elementary formal systems and formal languages-simple formal systems. *Memoirs of the Faculty of Science, Kyushu University. Series A, Mathematics*, Vol. 24, No. 1, pp. 47–75, 1970.

[2] 有川節夫, 西野哲朗, 石坂裕毅. 形式言語の理論. 情報科学コアカリキュラム講座. 丸善, 1999.

[3] Setsuo Arikawa, Takeshi Shinohara, and Akihiro Yamamoto. Learning elementary formal system. *Theoretical Computer Science*, Vol. 95, No. 1, pp. 97–113, 1992.

[4] G. Rozenberg and A. Salomaa, editors. *Handbook of Formal Languages:Word, language, grammar*, Vol. 1. Springer-Verlag, 1997.

[5] R. M. Smullyan. *Theory of Formal Systems*. Princeton Univ. Press, 1961.