非終端記号を導入した基本形式体系の言語記述力について

On the Expressive Power of Elementary Formal System with Nonterminals

小出智彦　　　篠原歩
Tomohiko Koide　　　Ayumi Shinohara

東北大学大学院 情報科学研究科
Graduate School of Information Science, Tohoku University

1　序論

基本形式体系 (Elementary Formal System, EFS) は Smullyan [4] 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．EFS は形式文法が持つ言語を生成する機能とオートマトンが持つ言語を受理する機能の 2 つを合わせ持つ強力な体系である．

本論文では以下のことについて述べる．まず，EFS に関する多くの論文の定義とは異なる空文字を扱うことが可能な EFS(消去可能 EFS) を定義し，それによって生じた問題を解決する．次に，EFS の部分クラスの 1 つである変数純粋 EFS [5] の言語記述力について述べる．最後に，非終端記号付き基本形式体系 (Elementary Formal System with Nonterminals, NEFS) を提案し，その言語記述力や言語の受理について述べる．

2　準備

Σ，X，Π を互いに交わらない集合とする．さらに Σ と Π は有限とする．Σ の要素を**終端記号**，X の要素を**変数**，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．集合 A に対し A^* を A 中の文字によって作られる文字列全体，A^+ を $A^* - \{\epsilon\}$ とする．ϵ は空文字を表す．

$(\Sigma \cup X)^*$ の要素を**項**または**パターン**と呼ぶ．変数を含まない項のことを**基礎項**と呼ぶ．Π 中の n 引数述語記号 p と n 個の項 τ_i $(1 \leq i \leq n)$ に対し $p(\tau_1, ..., \tau_n)$ を**アトム**と呼ぶ．n 個の項が全て基礎項の場合には**基礎アトム**と呼ぶ．節，空節 ($\Box$)，基礎節，ゴール節，代入は [1] に従って定義する．A，B_i $(1 \leq i \leq n)$ をアトムとするとき，$A \leftarrow B_1, ..., B_n$ という形の節を**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．Π は Σ と Γ から構成可能な確定節の有限集合で**公理**と呼ばれる．

定義 1. $T = (\Sigma, \Pi, \Gamma)$ を *EFS* とし，C を T の節とする．関係 $\Gamma \vdash C$ を次のように定義する．

- $C \in \Gamma$ ならば，$\Gamma \vdash C$.
- $\Gamma \vdash C$ ならば，任意の代入 θ に対して $\Gamma \vdash C\theta$.
- $\Gamma \vdash A \leftarrow B_1, ..., B_n$ と $\Gamma \vdash B_n \leftarrow$ に対して $\Gamma \vdash A \leftarrow B_1, ..., B_{n-1}$.

$\Gamma \vdash C$ ならば C は Γ から**証明可能**であるという．

定義 2. *EFS* $T = (\Sigma, \Pi, \Gamma)$ と n 引数の述語記号 $p \in \Pi$ に対し，$L(T, p)$ を次のように定義する．

$$L(T, p) = \{(\alpha_1, ..., \alpha_n) \in (\Sigma^*)^n \mid \Gamma \vdash p(\alpha_1, ..., \alpha_n) \leftarrow\}$$

$n = 1$ のとき，$L(T, p)$ は Σ 上の言語である．そのような S と p が存在するとき，$L \subseteq \Sigma^*$ は**基本形式体系言語**であるという．

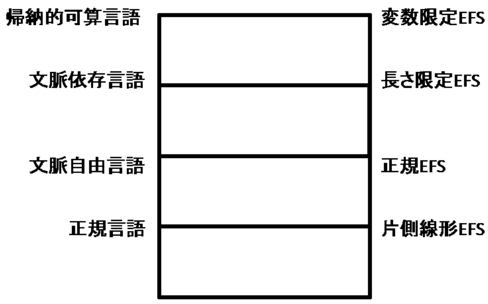

図 1　各 EFS 言語とチョムスキー階層の各言語との対応

Arikawa ら [2] はいくつかの部分クラスを提案し，その言語記述力が図 1 のようになることを示した．ただし，長さ限定 EFS 言語と文脈依存言語の等価性については制限がついていた．

2 つの節 C，D に対して，$C = D\theta$，$C\theta' = D$ なる代入 θ と θ' が存在するとき，D は C の**異種**であるという．

定義 3. T を *EFS* とし，G をゴール節とする．次の条件を満たす 3 項組 (G_i, θ_i, C_i) $(i = 0, 1, ...)$ の列を G からの導出と呼ぶ．

- G_i はゴール節，θ_i は代入，C_i は S の公理の異種である．また，$G_0 = G$.
- $i \neq j$ なる任意の i，j に対し $v(C_i) \cap v(C_j) = \emptyset$．また，任意の i に対し $v(C_i) \cap v(G_i) = \emptyset$.
- G_i が $\leftarrow A_1, ..., A_k$，A_m が R によって選ばれたアトムならば，C_i は $A \leftarrow B_1, ..., B_q$，$\theta_i$ は A と A_m の単一化代入，そして G_{i+1} は次のようになる．

$$(\leftarrow A_1, ..., A_{m-1}, B_1, ..., B_q, A_{m+1}, ..., A_k)\theta_i$$

G_{i+1} を G_i と C_i の θ_i による**融合節**と呼ぶ．

空節で終わる有限な導出を**反駁**と呼ぶ．

3　EFS の言語記述力

アトム A に対して，A 中に現れる全ての終端記号の集合を $t(A)$ で表す．また，パターン π における集合 S の要素の出現回数を $o_S(\pi)$ で表す．さらに，アトム $p(\pi_1, ..., \pi_n)$ に対して $o_S(p(\pi_1, ..., \pi_n)) = o_S(\pi_1) + ... + o_S(\pi_n)$ と定義する．

空代入を許した消去可能 EFS における節が長さ限定であることの必要十分条件は通常の空代入を許さない EFS とは異なり次のようになる．

定理 1. 節 $A \leftarrow B_1, ..., B_n$ に対し $T = t(A) \cup t(B_1) \cup ... \cup t(B_n)$ とするとき，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，

$$o_T(A) \geq o_T(B_1) + ... + o_T(B_n)$$

かつ，全ての変数 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．

次の定理は長さ限定 EFS 言語と文脈依存言語が厳密に等価であることを示している．

定理 2. 言語 $L \subseteq \Sigma^*$ が文脈依存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 *EFS* T が存在することである．

また，変数純粋 EFS [5] と並列多重文脈自由言語 [3] について次の定理が成り立つ．

定理 3. 言語 $L \subseteq \Sigma^*$ が並列多重文脈自由言語であるための必要十分条件は，$L = L(T, p)$ なる変数純粋 *EFS* T が存在することである．

4　NEFS の言語記述力

まず，EFS に非終端記号を導入した非終端記号付き基本形式体系を定義する．N を Σ，Π，X とは交わらない有限集合とする．N の要素を**非終端記号**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 と Σ と Π から構成可能な確定節の有限集合である．代入，証明可能性，$L(T, p)$，部分クラスついては EFS と同様に定義する．

非終端記号を導入したことにより形式文法やオートマトンを簡単に NEFS に変換することができる．例えばオートマトンを NEFS で模倣しようとする場合，テープアルファベットやヘッドを非終端記号で表現することにより受理する言語に影響を与えることなくテープの状況を NEFS で再現することができる．

非終端記号を導入したことにより言語記述力は高まっているように見えるが，各 NEFS の実際の言語記述力は次の通りである．

定理 4.
1. $L \in \Sigma^*$ が帰納的可算言語であることの必要十分条件は，$L = L(T, p)$ なる変数限定 *NEFS* T が存在することである．
2. $L \in \Sigma^*$ が文脈自由言語であることの必要十分条件は，$L = L(T, p)$ なる正規 *NEFS* T が存在することである．
3. $L \in \Sigma^*$ が正規言語であることの必要十分条件は，$L = L(T, p)$ なる片側線形 *NEFS* T が存在することである．

この結果を表 1 にまとめた．

NEFS $T = (N, \Sigma, \Pi, \Gamma)$ に対し，N と Σ と Π から構成可能な全ての非終端記号付き基礎アトムの集合を $B(T)$ と書く．さらに，$PS(T)$ と $RS(T)$ を次のように定義する．

表 1　チョムスキー階層の各言語と各 NEFS 言語の対応

	EFS	NEFS
帰納的可算	変数限定	変数限定
文脈自由	正規	正規
正規	片側線形	片側線形

$$PS(T) = \{A \in B(T) \mid \Gamma \vdash A \leftarrow\}$$
$$RS(T) = \{A \in B(T) \mid \leftarrow A \text{ からの反駁が存在する } \}$$

導出手続きについては EFS と同様に定義する．次の定理は NEFS も融合原理を用いることで言語の受理機構として機能することを示している．

定理 5. T が変数限定 *NEFS* ならば，次の等式が成り立つ．

$$PS(T) = RS(T)$$

5　まとめと今後の課題

本論文では消去可能 EFS における長さ限定 EFS の必要十分条件し，全ての長さ限定 EFS が変数限定 EFS であることを示した．また，長さ限定 EFS 言語が文脈依存言語と厳密に等価であること，変数純粋 EFS 言語が並列多重文脈自由言語と等価であることを示した．さらに，NEFS を定義し，その言語記述力と融合原理による言語の受理が可能であることを示した．

本論文では変数純粋 EFS を取り上げたものの，EFS の部分クラスは他にも数多く提案されている．それらの EFS で定義される言語がどのクラスに属するか調べるのが今後の課題である．

参考文献

[1] 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.

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

[3] Hiroyuki Seki, Takashi Matsumura, Mamoru Fujii, and Tadao Kasami. On multiple context-free grammars. *Theoretical Computer Science*, Vol. 88, No. 2, pp. 191–229, 1991.

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

[5] Noriko Sugimoto and Hiroki Ishizaka. Generating languages by a derivation procedure for elementary formal systems. *Information Processing Letters*, Vol. 69, No. 4, pp. 161–166, 1999.