数理解析研究所講究録
1489巻 2006年 202-208

拡張Dyck言語によるTALsの特徴づけ

電気通信大学電気通信学研究科情報工学専攻　松原 俊一 (Shunichi MATSUBARA)

Department of Computer science,TheUniversity of Electro Communication

概要

TALsは文脈自由言語よりも強力なもつ生成能力を要する言語のクラスであり，自然言語の形式化として近年注目を集めている．この論文では，任意のTALが認識可能集合と拡張Dyck言語によって特徴づけられることを示す．

1 まえがき

Tree Adjoining Languages(TALs)[5] は，近年，自然言語のクラスとして注目を集めており，多くの研究が盛んに行われている．この背景にあるのは，自然言語の形式化には文脈自由よりもやや強力な生成能力の文法が必要であろう，という主張である．この主張は，現在，有力なものになっており，そうした能力をもつ文法として導入されたものに Tree AdjoiningGrammars(TAGs)[5] やSpine文法[3]といったものがある．TALsはそれらの文法が生成する文字列言語のクラスである．

自然言語の形式的な文法として別々に導入された Head Grammars, TAGs, Combinatory Categorial Grammars, linear Indexed Grammars の４つの文法の生成能力が一致するという結果[6]は，TALsが自然言語のクラスなのではないかという推論に強い根拠を与える興味深い事実である．また，[3]によって導入されたSpine文法の生成する言語も，TALsに属することが示されている．Spine文法は，簡潔な書き換え規則によって表現できるということも知られている[3].

以下では，Dyck言語を木言語へと一般化した拡張Dyck言語Dを導入し，TALsを特徴づける．本論文の主定理は次である．任意のTAL Lに対して，認識可能集合Rが存在して $\mathrm{L} = yield(\mathrm{R} \cap \mathrm{D})$ となる．

2 準備

N を非負整数全体からなる集合，N_+ を正の整数全体からなる集合とする．W を集合とするとき，W^* を W 上の自由モノイドとし，単位元を λ で表す．また，$W^+ = W^* - \{\lambda\}$ とする．空集合を $\emptyset$ で表す．

2.1 木

木の台集合 $D \subseteq N_+^*$ を次の条件を満たす空でない有限集合であると定義する．$d \in D$ に対し,d' が d の前部分語ならば，$d' \in D$ であり，かつ $d \cdot i \in D$ に対し,$i \in N_+$, $1 \leq j \leq i$ ならば，$d \cdot j \in D$.

木の台集合 D の元を**節点**，節点 λ を**根**という．また，D の元 d，非負整数 i に対して，$d \cdot i \in D$ であるならば $d \cdot i$ を d の**子**，d を $d \cdot i$ の**親**とよび，子のない節点を**葉**とよぶ．**節点 d_1 から節点 d_2 への道** $path(d_1, d_2)$ とは $path(d_1, d_2) = \{d \in D | d_2$が d の前部分語であり，かつ d が d_1の前部分語$\}$ で定義される節点の有限集合のことである．$path(d_1, d_2)$ の元を**道 $path(d_1, d_2)$ 上の節点**という．

階層アルファベットとは (Ψ, r) のことである．ここで，Ψ は有限集合，r は $r : \Psi \to N$ で定義される関数である．Ψ の元 a に対して，$r(a)$ を a の**階層**とよぶ．また，$n \in N_+$ に対し，$\Psi_n = r^{-1}(n)$ と定義する．さらに，階層アルファベットは常に特別な記号 x と ε を含むものとする．x は変数であり，$r(x) = 0$ で以下で定義する木の代入が許される記号．ε は $r(x) = 0$ の空語を表す記号とする．

主要部指示付階層アルファベットとは (Ψ, r, h) のことで，階層アルファベット (Ψ, r) 上に部分関数 h を定義したものである．h は Ψ から N_+ への部分関数であり，Ψ の元 a に対して，$r(a) = 0$ のときは定義されず，$r(a) \geq 1$ のとき $1 \leq h(a) \leq r(a)$ を満たす．Ψ の元 a に対して，$h(a)$ が定義されているならば $h(a)$ を a の**主要部指示**とよぶ．

以下では階層アルファベット (Ψ, r) を単に階層アルファベット Ψ，主要部指示階層アルファベット (Ψ, r, h) を主要部指示アルファベット Ψ と表現する．また，Ψ は主要部指示付階層アルファベットを表すものとする．

Ψ 上の**木** を関数 $\alpha : D \to \Psi$ で定義する．木 α の

定義域を D_α で表し，$\alpha(d)$ を節点 d の**ラベル**とよぶ．また，α を Ψ 上の木，d を D_α に属する葉ではない節点とするとき，$d \cdot h(\alpha(d))$ を d の**主要部節点**とよぶ．

Ψ 上の木 α, β，D_α の節点 d，$l \in N_+^*$ に対し，木に関する 2 つの演算を次で定義する．

- $\alpha/d = \{(d', a) | d' \in N_+^*, (d \cdot d', a) \in \alpha\}$
- $\alpha(d \leftarrow \beta) = \{(d', a) | (d', a) \in \alpha, d$ は d' の前部分語ではない $\} \bigcup \{(d \cdot d', a) | (d', a) \in \beta\}$

主要部指示付階層アルファベット上の木 α とし，**spine** という概念を定義する．spine とは根から葉への道のうち，その道に属する葉以外の節点 d に対して，d の主要部節点もまたその道上の節点であるようなもののことである．木 α の spine を spine(α) と表す．また，spine(α) 上の節点のうち葉であるものを **footnode** といい，$ft(\alpha)$ と書く．

spine(α) の節点を根から順に $d_0, d_1, \cdots, d_l$ とする．このときラベルの列 $\alpha(d_0)\alpha(d_1)\cdots\alpha(d_l)$ を **spine ラベル列**とよぶ．

木 α の節点 d が持つ子の数を $\deg_\alpha(d)$ で表し，$\deg_\alpha(d) = max\{i \in N_+ | d \cdot i \in D_\alpha\}$ と定義する．$\deg_\alpha(d)$ のことを節点 d の**次数**という．

α を Ψ 上の木であるとする．このとき D_α の元 d が $\deg_\alpha(d) = r(\alpha(d))$ を満たし，かつ，$d \neq ft(\alpha)$ ならば $\alpha(d) \neq x$ を満たすとき木 α を **整形木**と呼び，Ψ 上の**整形木全体からなる集合** Ψ^T を $\Psi^T = \{\alpha | \alpha$ は Ψ 上の整形木$\}$ で定義する．また，$\alpha \in \Psi^T$ の footnode が x であるとき α のことを $\alpha(x)$ と書くこともある．

木を表すのに，節点のラベルを根から順にポーランド記法で並べた文字列を用いることがある．これを**文字列表現**とよぶ．木 α の文字列表現は，根のラベル a の階層を i，α/i の文字列表現を α_i，としたとき $a\alpha_1\alpha_2\cdots\alpha_n$ で表される文字列として定義する．ここで，文字列表現は木自体を現すときに用いることに注意する．

関数 *yield* : $\Psi^T \to \Psi^*$ を帰納的に次で定義する．$D_\alpha = \{\lambda\}$，$\alpha(\lambda) \neq \varepsilon$ のとき $yield(\alpha) = \alpha(\lambda)$．$D_\alpha = \{\lambda\}$，$\alpha(\lambda) = \varepsilon$ のとき $yield(\alpha) = \lambda$．$D_\alpha \neq \{\lambda\}$ のとき $yield(\alpha) = yield(\alpha/1)yield(\alpha/2)\cdots yield(\alpha/r(\alpha(\lambda)))$

$\alpha, \beta \in \Psi^T$ とするとき，β を α に**代入**した結果を $\alpha[\beta]$ で表し，$\alpha[\beta] = \{(d, a) \in \alpha | a \neq x\} \bigcup \{(d' \cdot d, a) | (d', x) \in \alpha, (d, a) \in \beta\}$ で定義する．

2.2 Spine 文法

Spine 文法は，[3] によって導入され，生成する文字列言語のクラスが TALs に一致することなどが証明されている．定義 2.1 の Spine 文法は，[3] で導入されたものとはやや異なるが，生成能力に関して，[3] で導入された Spine 文法と同等であることは明らかである．

定義 2.1 主要部指示付階層アルファベット Δ 上の **Spine 文法** とは次で定義される *4* 項組 $G_S = (\Gamma, \Delta, P, S)$ のこと．ここで，Γ は $\Gamma = \Gamma_0 \bigcup \Gamma_1, \Gamma \bigcap \Delta = \emptyset$ であるような主要部指示付階層アルファベットで，その元を**非終端記号**という．Δ の元を**終端記号**という．S は Γ の元で**開始記号**という．P は次で示される形の書き換え規則の集合，

1. $A \to \alpha, \alpha \in (\Delta - \{x\})^T$
2. $Bx \to \beta(x), \beta \in \Delta^T$

$G = (\Gamma, \Delta, P, S)$ を Spine 文法とする．G の d における $\pi \in P$ の適用による **1 ステップ導出** $\overset{\pi}{\underset{G}{\Rightarrow}}$ とは，$(\Gamma \bigcup \Delta)^T$ 上の関係であり，$\pi = \phi_1 \to \phi_2$ とすると次で定義される．$\alpha \in (\Gamma \bigcup \Delta)^T$ と $\beta \in (\Gamma \bigcup \Delta)^T$ に対して，$\psi \in (\Gamma \bigcup \Delta)^T$ が存在し，$\alpha/d = \phi_1[\psi]$，$\beta = \alpha(d \leftarrow \phi_2[\psi])$ であるとき，かつそのときに限り $\alpha \overset{\pi}{\underset{G}{\Rightarrow}} \beta$．ただし，G や π が明らかなときは，これを省略する．$n \geq 0$ に対する G の $\pi_1\pi_2\cdots\pi_n \in P^*$ の適用による **n ステップ導出** $\overset{\pi_1\pi_2\cdots\pi_n}{\underset{G}{\Longrightarrow}}$ とは，$(\Gamma \bigcup \Delta)^T$ 上の関係であり，次で定義される．$\alpha, \beta \in (\Gamma \bigcup \Delta)^T$ に対して $\delta_0, \delta_1, \cdots, \delta_n \in (\Gamma \bigcup \Delta)^T$ が存在して，$\alpha = \delta_0 \overset{\pi_1}{\underset{G}{\Rightarrow}} \delta_1, \delta_1 \overset{\pi_2}{\underset{G}{\Rightarrow}} \delta_2, \cdots, \delta_{n-1} \overset{\pi_2}{\underset{G}{\Rightarrow}} \delta_n = \beta$ であるとき，かつそのときに限って $\alpha \overset{\pi_1\pi_2\cdots\pi_n}{\underset{G}{\Longrightarrow}} \beta$．ただし，書き換え規則の列を書かずに $\overset{n}{\Rightarrow}$ と書くこともある．n ステップ導出の反射推移閉包を $\overset{*}{\Rightarrow}$ で表し，ただ単に**導出** とよぶ．

$G_S = (\Gamma, \Delta, P, S)$ を Spine 文法とする．集合 $T(G_S) = \{\alpha \in \Delta^T | S \overset{*}{\Rightarrow} \alpha\}$ を G_S の生成する **Spine 木言語**という．集合 $L(G_S) = \{w \in \Delta_0^* | \exists \alpha \in \Delta^T, S \overset{*}{\Rightarrow} \alpha, yield(\alpha) = w\}$ を G_S の生成する **Spine 言語**という．

Spine 文法 G_1, G_2 に対して，$T(G_1) = T(G_2)$ であるならば，G_1 と G_2 は**同値** であるといい．$L(G_1) = L(G_2)$ であるならば，G_1 と G_2 は**弱同値** であるという．

2.3 正則系

正則系は，[1] によって導入され，認識可能集合を生成することや，木オートマトン ([1]) によって受理されることなどが知られている．また，文字列上の正則文法の木への一般化と捉えることもできる．正則系の定義には様々な形があるが，この論文では Spine 文法に制限を加えたものとして次で定義する．

定義 2.2 主要部指示付階層アルファベット Δ 上の**正則系** とは，*4* 項組 $G_R = (\Gamma, \Delta, P, S)$ のことで，*Spine* 文法 のうち書き換え規則が次の形のもののみからなるもの．

- $A \to \alpha, A \in \Gamma_0, \alpha \in (\Delta - \{x\})^T$

2.4 強標準形

この論文では，Spine 文法によって生成される言語のうち，TALs と弱同値である Spine 言語に焦点がある．Spine 言語は主要部指示付階層アルファベットの階層 0 の元のみからなる文字列言語であることから，以下では階層 1 以上の元が τ, θ の 2 つのみからなる主要部指示付階層アルファベット上で議論する．ここで，$r(\tau) = r(\theta) = 2$，$h(\tau) = 2$，$h(\theta) = 1$ とし，Σ をこのように約束した主要部指示付階層アルファベットとする．このとき注意すべきことは，Σ 上で定義される Spine 文法は，任意の Spine 木言語を生成することはできないが，生成する文字列言語に関しては一般性を失っていないことである．すなわち，Σ 上の Spine 文法に関する議論は任意の Spine 言語に関するものである．

任意の Spine 言語は，[定義 2.3] に示すこの単純な形で定義できることが [3] で証明されている．[定義 2.3] の形は [3] の定義とはわずかに異なるが，両者が同じクラスの文字列言語を生成することは明らかである．

定義 2.3 *Spine* 文法 $G_S = (\Gamma, \Sigma, P, S)$ のうち全ての書き換え規則が以下のいずれかの形で表せるとき，G_S は**強標準形**であるという．

1. $A \to a$，$A \in \Gamma_0$，$a \in \Sigma_0$

2. $A \to BC$，$A, C \in \Gamma_0$，$B \in \Gamma_1$

3. $B_1 x \to B_2 B_3 x$，$B_1, B_2, B_3 \in \Gamma_1$

4. $Bx \to \tau A x$，$A \in \Gamma_0$，$B \in \Gamma_1$

5. $Bx \to \theta x A$，$A \in \Gamma_0$，$B \in \Gamma_1$

これ以降，L は TAL を，R は認識可能集合を表すものとする．

3 拡張 Dyck 言語 D

Dyck 言語とは，数種類の括弧が整合しているような文脈自由言語である．任意の文脈自由言語が Dyck 言語と適当な正則言語との共通集合の準同型写像になっているという特徴付けが [2] によって行われている．本論文における TALs の特徴付けは，この特徴づけの木への一般化であるともいえる．また，[2] の文脈自由言語の特徴づけの発展的な研究結果として文脈自由普遍文法 ([4]) が導入されている．

この節では，まず，文字列における括弧の整合の概念を木の上に拡張する．そして，文字列言語として考えられた Dyck 言語を木言語に拡張した拡張 Dyck 言語を新しく導入し，この言語が Spine 木言語と同値であることを示す．

$\{e, \bar{e}, f, \bar{f}\}$ を Σ と互いに素な指示付階層アルファベットとする．ここで，$r(e) = r(\bar{e}) = r(f) = r(\bar{f}) = 1$．この $e, \bar{e}$ と $f, \bar{f}$ はそれぞれ対になる括弧の組を意味している．指示付階層アルファベット Ω を $\Omega = \Sigma_0 \bigcup \{e, \bar{e}, f, \bar{f}\}$ で定義する．

定義 3.1 $S \notin \Omega$ とする．$(\Omega \bigcup \{S\})^T$ から $(\Omega \bigcup \{S\})^T$ への関係 ~ を図 *1* で定義する．ただし，*a* は Ω_0 の任意の元を表す．

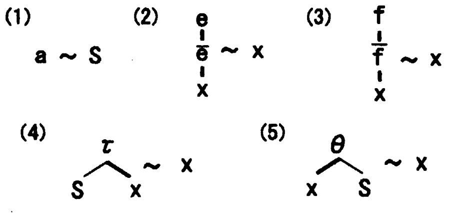

図 1: 還元規則

関係 ~ の元を**還元規則** という．

次に，木のある部分に対して還元規則を適用するために，関係 ≈ を定義する．

定義 3.2 $(\Omega \bigcup \{S\})^T$ 上の関係 ≈ を ***1* ステップ還元**といい，次で定義する．

- $\alpha,\beta \in (\Omega \bigcup \{S\})^T$ とする．このとき $\alpha \approx \beta$ となるのは，$\alpha/d = \phi_1[\psi]$，$\phi_1 \sim \phi_2$，$\beta = \alpha(d \leftarrow \phi_2[\psi])$ を満たすような D_α の元 d と $\Omega \bigcup \{S\}$ 上の整形木 ϕ_1, ϕ_2, ψ が存在するとき，かつそのときに限る．

定義 3.3 $(\Omega \bigcup \{S\})^T$ 上の関係 $\overset{n}{\approx}$ を **n ステップ還元**，あるいはただ単に**還元**といい，次で定義する．

- $\alpha,\beta \in (\Omega \bigcup \{S\})^T$，$n$ を非負整数，i を n より小さい非負整数とする．このとき $\alpha \overset{n}{\approx} \beta$ となるのは，$\delta_0 = \alpha$，$\delta_n = \beta$，$\delta_i \approx \delta_{i+1}$ を満たすような $\Omega \bigcup \{S\}$ 上の整形木 $\delta_0, \delta_1, \cdots, \delta_n$ が存在するとき，かつそのときに限る．

図 1 の還元規則は，次の動機によって導入されたものである．(2) と (3) は 1 組の括弧の対が正しく現れたとき還元を行う規則である．(4) と (5) は，部分木 δ に対して，$spine(\delta)$ 上の節点から主要部指示がない方に分岐した部分木 δ' の全ての括弧の整合性が確認されたときに，δ' と $spine(\delta)$ 上の τ あるいは θ との消去を行う規則である．また，(4), (5) は括弧の整合の概念を木へ拡張するための規則ともいえる．

定義 3.4 拡張 *Dyck* 言語 D を次で定義する．

- $\mathrm{D} = \{\alpha \in \Omega^T | \alpha \overset{*}{\approx} S\}$

本節の以下では，D が TALs に属することを示す．まず，次を定義する．

定義 3.5 $\beta(x) \in \Psi^T$ とするとき，次の形の書き換え規則を**空記号規則**とよぶ．

- $x \rightarrow \beta(x)$

補題 3.1 *Spine* 文法の書き換え規則の定義に空記号規則を加えた文法の生成能力は *Spine* 文法と同じである．換言すれば，*Spine* 文法の定義を拡張し空記号規則を許しても，その生成能力は変わらない．

(証明) 証明内では，拡張する前の Spine 文法を単に Spine 文法と呼び，空記号規則を許すように拡張した Spine 文法を拡張した Spine 文法と呼ぶ．

Spine 文法の生成するクラスが，拡張した Spine 文法のクラスに含まれることは明らかである．よって以下では，拡張した Spine 文法のクラスが Spine 文法のクラスに属すことを示す．

拡張した Spine 文法 $\mathrm{G} = (\Gamma, \Delta, P_1 \bigcup P_2, S)$ から Spine 文法 $\mathrm{G}' = (\Gamma, \Delta, P_1 \bigcup P_2', S)$ を構成する．ただし，P_1 は空記号規則ではない書き換え規則の有限集合とし，P_2 は空記号規則のみからなる書き換え規則の有限集合とする．ここで，P_2' は次で構成される．

```
begin
 TEMP1 := P2;
 while TEMP1 ≠ ∅
  begin
   x → β(x) ∈ TEMP1 を 1 つ TEMP1 から
   取り除く；
   TEMP2 := P1;
   while TEMP2 ≠ ∅
    begin
     ζ → η ∈ TEMP2 を 1 つ TEMP2 から
     取り除く；
     TEMP3 := Dη;
     while TEMP3 ≠ ∅
      begin
       d ∈ TEMP3 を 1 つ TEMP3 から
       取り除く；
       ζ → η(d ← β[η/d]) を P2' に
       加える；
      end
    end
  end
end
```

$T(\mathrm{G}) = T(\mathrm{G}')$ であることは構成法より明らか．

補題 3.2 拡張 *Dyck* 言語 *D* は *TALs* に属す．

(証明) D を生成する Spine 文法 $\mathrm{G_D}$ は $\mathrm{G_D} = (\{S\}, \Omega, P_\mathrm{D}, S)$ として構成できる．ただし，$P_\mathrm{D} = \{\alpha \rightarrow \beta | \beta \sim \alpha\}$．Spine 文法 $\mathrm{G_D}$ が D を生成することは構成法より明らか．

4 TALs の特徴付け

この節では，任意の L に対して，$\mathrm{L} = yield(\mathrm{R} \bigcap \mathrm{D})$ を満たすような R が存在する，という主定理を証明し，TALs を特徴付ける．

4.1 標準導出

この小節では，定理 4.2 の証明の準備として Spine 文法の標準導出を導入する．

定義 4.1 $\delta_0, \delta_1, \cdots, \delta_n \in (\Gamma \bigcup \Omega)^T$，$n \geq 0$ とする．*Spine* 文法 $\mathrm{G_S} = (\Gamma, \Omega, P, S)$ の導出 $\delta_0 \overset{\pi_1\pi_2\cdots\pi_n}{\Longrightarrow} \delta_n$ が次を満たすとき，この導出を**標準導出**という．ただ

し，$0 \le i \le n$ に対して，$\delta_i \overset{\pi_i}{\Longrightarrow} \delta_{i+1}$ を d における 1 ステップ導出とする．

- d の真の前部分語 d' に対して，
  1. d' の子 d'' が d' の主要部節点ではなく，d の前部分語でもないならば $\delta_i/d'' \in \Omega^T$
  2. d' のラベルは終端記号．

定義 4.1 の (2) を満たす導出は，**OI derivation** ([7]) とよばれ，任意の TAL の全ての語がこの導出によって導かれることが示されている ([3],[7])．また，[1] で示されている木の演算に関するいくつかの補題より，次の補題 4.1 が成り立つことは明らかである．

補題 4.1 Σ_0 上の文字列 w が Σ 上の *TALs* に属するならば $w = yield(\alpha)$ となる $\alpha \in \Sigma^T$ を導く *Spine* 文法の標準導出が存在する．

4.2 主定理

以下の証明中に現れる導出は，全て標準導出であるとする．

定理 4.1 任意の *TAL* L に対して，認識可能集合 R が存在して，$\mathrm{L} = yield(\mathrm{R} \cap \mathrm{D})$ を満たす．

(証明)　まず，強標準形の Spine 文法 $G_S = (\{C_1, C_2, \cdots, C_l\}, \Sigma, P_S, C_1)$ から正則系 $G_R = (\{S\}, \Omega, P_R, S)$ を構成する．ここで，g を次で定義される $\Omega^T \times (\Omega \cup \{S\})^T$ 上の部分関数とし，$P_R = \{g(\pi) | \pi \in P_S\}$ とする．

- $g(C_i \to a) = S \to \bar{e}^i \bar{f} a$
- $g(C_i \to C_j C_k) = S \to \bar{e}^i \bar{f} f e^k f e^j S$
- $g(C_i x \to \tau C_j x) = S \to \bar{e}^i \bar{f} \tau f e^j S S$
- $g(C_i x \to \theta x C_j) = S \to \bar{e}^i \bar{f} \theta S f e^j S$
- $g(C_i x \to C_j C_k x) = S \to \bar{e}^i \bar{f} f e^k f e^j S$

この定理を証明するためには，$n \ge 0$ とするとき，Σ_0 上の文字列 w, u, v に対して，次の 2 つの命題が成り立つことを示せばよい．

1. $C_i \overset{\pi_1\pi_2\cdots\pi_n}{\underset{G_S}{\Longrightarrow}} \alpha$, $w = yield(\alpha)$ を満たす $\alpha \in \Sigma^T$ が存在するための必要十分条件は $S \overset{g(\pi_1)g(\pi_2)\cdots g(\pi_n)}{\underset{G_R}{\Longrightarrow}} \gamma$, $fe^i\gamma \overset{*}{\approx} S$, $yield(\gamma) = w$ を満たす $\gamma \in \Omega^T$ が存在することである．
2. $C_i x \overset{\pi_1\pi_2\cdots\pi_n}{\underset{G_S}{\Longrightarrow}} \beta$, $yield(\beta) = uxv$ を満たす $\beta(x) \in \Sigma^T$ が存在するための必要十分条件は $S \overset{g(\pi_1)g(\pi_2)\cdots g(\pi_n)}{\underset{G_R}{\Longrightarrow}} \delta[S]$, $fe^i\delta \overset{*}{\approx} x$, $yield(\delta) = uxv$ を満たす $\delta(x) \in \Omega^T$ が存在することである．

(1) と (2) を導出の長さ n 関する帰納法によって同時に示す．

(必要条件の証明) (1) の基底 長さ 0 のときは前提が偽なので成立．長さ 1 のとき導出は $C_i \to a, a \in \Sigma$ の形の書き換え規則 1 回の適用に限られる．すなわち $S \overset{\pi}{\underset{G_S}{\Rightarrow}} \alpha = a$, $w = yield(\alpha) = a$．このとき，P_R の構成法より，$S \to \bar{e}^i \bar{f} a \in P_R$．したがって $S \overset{g(\pi)}{\underset{G_R}{\Rightarrow}} \bar{e}^i \bar{f} a = \gamma$ $fe^i\gamma \overset{*}{\approx} S$, $yield(\gamma) = a = w$ を満たす $\gamma \in \Omega^T$ が存在し，成立．長さ 2,3 のときも前提は偽になるので成立．

(2) の基底 長さ 0,1, のとき前提は偽．よって成立．

(1) の帰納ステップ $n > 3$ とする．Σ_0 上の文字列 w に対して $C_i \overset{\pi_1\pi_2\cdots\pi_n}{\underset{G_S}{\Longrightarrow}} \alpha, w = yield(\alpha)$ を満たすような $\alpha \in \Sigma^T$ が存在するとする．このとき導出は次のように表せる．$C_i \overset{\pi_1}{\underset{G_S}{\Rightarrow}} C_j C_k \overset{\pi_2\pi_3\cdots\pi_m}{\underset{G_S}{\Longrightarrow}} \beta_1[C_k] \overset{\pi_{m+1}\pi_{m+2}\cdots\pi_n}{\underset{G_S}{\Longrightarrow}} \beta_1[\alpha_1] = \alpha$．ここで，$1 < m < n$, $\pi_1 = C_i \to C_i C_j$, $\alpha \in \Sigma^T, \beta_1(x) \in \Sigma^T$, $C_j x \overset{\pi_2\pi_3\cdots\pi_m}{\underset{G_S}{\Longrightarrow}} \beta_1$, $C_k \overset{\pi_{m+1}\pi_{m+2}\cdots\pi_n}{\underset{G_S}{\Longrightarrow}} \alpha_1$, $yield(\beta_1) = uxv, yield(\alpha_1) = w', yield(\alpha) = uw'v = w$．とする．$C_j x \overset{\pi_2\pi_3\cdots\pi_m}{\underset{G_S}{\Longrightarrow}} \beta_1$ は長さが n より小さい導出なので帰納法の仮定より，$S \overset{g(\pi_2)g(\pi_3)\cdots g(\pi_m)}{\underset{G_R}{\Longrightarrow}} \delta_1[S], fe^j\delta_1 \overset{*}{\approx} x, yield(\delta_1) = uxv$ を満たす $\delta_1 \in \Omega^T$ が存在する．同様に，$C_k \overset{\pi_{m+1}\pi_{m+2}\cdots\pi_n}{\underset{G_S}{\Longrightarrow}} \alpha_1$ に対しても帰納法の仮定がいえるので，$S \overset{g(\pi_{m+1})g(\pi_{m+2})\cdots g(\pi_n)}{\underset{G_R}{\Longrightarrow}} \gamma_1, fe^k\gamma_1 \overset{*}{\approx} S, yield(\gamma_1) = w'$ を満たす $\gamma_1 \in \Omega^T$ が存在する．一方で，$\pi_1 \in P_S$ であるから P_R の構成法より，$g(\pi_1) = S \to \bar{e}^i \bar{f} f e^k f e^j S \in P_R$ である．したがって，$S \overset{g(\pi_1)}{\underset{G_R}{\Longrightarrow}} \bar{e}^i \bar{f} f e^k f e^j S \overset{g(\pi_2)g(\pi_3)\cdots g(\pi_m)}{\underset{G_R}{\Longrightarrow}} \bar{e}^i \bar{f} f e^k f e^j \delta_1[S] \overset{g(\pi_{m+1})g(\pi_{m+2})\cdots g(\pi_n)}{\underset{G_R}{\Longrightarrow}} \bar{e}^i \bar{f} f e^k f e^j \delta_1[\gamma_1] = \gamma$, $fe^i\gamma = fe^i\bar{e}^i \bar{f} f e^k f e^j \delta_1[\gamma_1] \overset{*}{\approx} fe^k f e^j \delta_1[\gamma_1] \overset{*}{\approx} fe^k\gamma_1 \approx S$, $yield(\gamma) = yield(\bar{e}^i \bar{f} f e^k f e^j \delta_1[\gamma_1]) = uw'v = w$ を満たす $\gamma \in \Omega^T$ が存在する．よって，成立．

(2) の帰納ステップ $n > 1$ とする．Σ_0 上の文字列 u, v に対して $C_i x \overset{\pi_1\pi_2\cdots\pi_n}{\underset{G_S}{\Longrightarrow}} \beta, uxv = yield(\beta)$ を満たすような $\beta(x) \in \Sigma^T$ が存在するとする．このとき π_1 として $C_i x \to \tau C_j x$, $C_i x \to \theta x C_j$, $C_i x \to C_j C_k x$ の 3 つの形の書き換え規則が考えられる．

$\pi_1 = C_i x \to \tau C_j x$ のときと $\pi_1 = C_i x \to \theta x C_j$ のとき対称的であるから，どちらか一方を示せばよいことに注意する．

$\pi_1 = C_i x \to \tau C_j x$ のときを考える．このとき導出は $C_i x \overset{\pi_1}{\underset{G_S}{\Rightarrow}} \tau C_j x \overset{\pi_2\pi_3\cdots\pi_n}{\underset{G_S}{\Longrightarrow}} \tau\alpha_1 x = \beta$ と表される．ここで，$\alpha_1 \in \Sigma^T$ は $C_j \overset{\pi_2\pi_3\cdots\pi_n}{\underset{G_S}{\Longrightarrow}} \alpha_1$，$yield(\alpha_1) = u$ を満たし，$v = \lambda$，$yield(\beta) = uxv$ である．すると，帰納法の仮定より，$S \overset{g(\pi_2)g(\pi_3)\cdots g(\pi_n)}{\underset{G_R}{\Longrightarrow}} \gamma_1, fe^j\gamma_1 \overset{*}{\approx} S, yield(\gamma_1) = v$ を満たす $\gamma_1 \in \Omega^T$ が存在する．一方で $\pi_1 \in P_S$ であるから P_R の構成法より，$g(\pi_1) = S \to \bar{e}^i\bar{f}\theta S fe^j S$ が P_S に属す．したがって，$S \overset{g(\pi_1)}{\underset{G_R}{\Longrightarrow}} \bar{e}^i\bar{f}\theta S fe^j S \overset{g(\pi_2)g(\pi_3)\cdots g(\pi_n)}{\underset{G_R}{\Longrightarrow}} \bar{e}^i\bar{f}\theta S fe^j\gamma_1 = \delta[S]$，$fe^i\delta = fe^i\bar{e}^i\bar{f}\theta x fe^j\gamma_1 \overset{*}{\approx} \theta x fe^j\gamma_1 \overset{*}{\approx} \theta x S \approx x$，$yield(\delta) = yield(\bar{e}^i\bar{f}\theta x fe^j\gamma_1) = yield(x)yield(\gamma_1) = xv = uxv$ を満たす $\delta(x) \in \Omega^T$ が存在する．よって，$\pi_1 = C_i \to \tau C_j x$ のとき成立．

次に，$\pi_1 = C_i x \to C_j C_k x$ のときを示す．このとき導出は，$C_i x \overset{\pi_1}{\underset{G_S}{\Rightarrow}} C_j C_k x \overset{\pi_2\pi_3\cdots\pi_m}{\underset{G_S}{\Longrightarrow}} \beta_1[C_k x] \overset{\pi_{m+1}\pi_{m+2}\cdots\pi_n}{\underset{G_S}{\Longrightarrow}} \beta_1[\beta_2] = \beta$ と表される．ここで，$1 < m < n$ であり，$\beta_1(x) \in \Sigma^T$ は $C_j x \overset{\pi_2\pi_3\cdots\pi_m}{\underset{G_S}{\Longrightarrow}} \beta_1$ を，$\beta_2(x) \in \Sigma^T$ は $C_k x \overset{\pi_{m+1}\pi_{m+2}\cdots\pi_n}{\underset{G_S}{\Longrightarrow}} \beta_2$ を満たす．また，$yield(\beta_1) = u_1xv_1$，$yield(\beta_2) = u_2xv_2$ とおくと，$yield(\beta) = u_1u_2xv_2v_1 = uxv$．と表される．すると，帰納法の仮定より，$S \overset{g(\pi_2)g(\pi_3)\cdots g(\pi_m)}{\underset{G_R}{\Longrightarrow}} \delta_1[S], fe^j\delta_1 \overset{*}{\approx} x, yield(\delta_1) = u_1xv_1$ を満たす $\delta_1(x) \in \Omega^T$ と，$C_k x \overset{\pi_{m+1}\pi_{m+2}\cdots\pi_n}{\underset{G_S}{\Longrightarrow}} \beta_2$ $S \overset{g(\pi_{m+1})g(\pi_{m+2})\cdots g(\pi_n)}{\underset{G_R}{\Longrightarrow}} \delta_2[S], fe^k\delta_2 \overset{*}{\approx} x, yield(\delta_2) = u_2xv_2$ を満たす $\delta_2(x) \in \Omega^T$ が存在する．一方で，$\pi_1 \in P_S$ であるから P_R の構成法より，$g(\pi_1) = S \to \bar{e}^i\bar{f}fe^kfe^jS \in P_R$ である．したがって，$S \overset{g(\pi_1)}{\underset{G_R}{\Longrightarrow}} \bar{e}^i\bar{f}fe^kfe^jS \overset{g(\pi_2)g(\pi_3)\cdots g(\pi_m)}{\underset{G_R}{\Longrightarrow}} \bar{e}^i\bar{f}fe^kfe^j\delta_1[S] \overset{g(\pi_{m+1})g(\pi_{m+2})\cdots g(\pi_n)}{\underset{G_R}{\Longrightarrow}} \bar{e}^i\bar{f}fe^kfe^j\delta_1[\delta_2[S]] = \delta[S]$，$fe^i\delta = fe^i\bar{e}^i\bar{f}fe^kfe^j\delta_1[\delta_2] \overset{*}{\approx} fe^kfe^j\delta_1[\delta_2] \overset{*}{\approx} fe^k\delta_2 \approx x$，$yield(\delta) = yield(\bar{e}^i\bar{f}fe^kfe^j\delta_1[\delta_2]) = u_1u_2xv_2v_1 = uxv$ を満たす $\delta(x) \in \Omega^T$ が存在する．よって，$\pi_1 = C_i x \to C_j C_k x$ のとき成立．

以上より，必要条件は成り立つ．

(十分条件の証明) (1) の基底 長さ 0 のとき前提は偽なので成立．Σ_0 上の文字列 w に対して，導出の長さが 1 のとき前提が成立するのは，導出が $S \overset{g(\pi_1)}{\underset{G_R}{\Longrightarrow}} \bar{e}^i\bar{f}a = \gamma$ のときのみである．ただし，$\pi_1 = C_i \to \bar{e}\bar{f}a, a \in \Sigma_0, a \neq x, fe^i\gamma \overset{*}{\approx} S, yield(\gamma) = a = w$ とする．すると，$g(\pi_1) \in P_R$ であるから P_R の構成法より $C_i \to a \in P_S$ である．したがって，$C_i \overset{C_i\to a}{\underset{G_S}{\Longrightarrow}} a = \alpha$，$yield(\alpha) = yield(a) = a = w$ を満たす $\alpha \in \Sigma^T$ が存在する．よって成立．

(2) の基底 長さ 0, 長さ 1 のときは前提が偽なので成立．

(1) の帰納ステップ $n > 1$ とする．Σ_0 上の文字列 w に対し，$S \overset{g(\pi_1)g(\pi_2)\cdots g(\pi_n)}{\underset{G_R}{\Longrightarrow}} \gamma, fe^i\gamma \overset{*}{\approx} S, yield(\gamma) = w$ を満たすような $\gamma \in \Omega^T$ が存在するとする．このとき $g(\pi_1)$ は，$S \to \bar{e}^i\bar{f}\tau fe^jSS$，$S \to \bar{e}^i\bar{f}\theta S fe^j S$，$S \to \bar{e}^i\bar{f}fe^kfe^jS$ のいずれかの形が考えられる．ただし，以下で示すように $S \to \bar{e}^i\bar{f}fe^kfe^jS$ の場合以外前提は偽となる．$g(\pi_1) = S \to \bar{e}^i\bar{f}\tau fe^jSS$ のとき前提が成立すると仮定すると，$S \overset{g(\pi_2)g(\pi_3)\cdots g(\pi_m)}{\Longrightarrow} \gamma_1, fe^j\gamma_1 \overset{*}{\approx} S$ を満たす $\gamma_1 \in \Omega^T$ と $S \overset{g(\pi_{m+1})g(\pi_{m+2})\cdots g(\pi_n)}{\Longrightarrow} \gamma_2, \gamma_2 \overset{*}{\approx} S$ を満たす $\gamma_2 \in \Omega^T$ が存在し，$S \overset{g(\pi_1)}{\underset{G_R}{\Longrightarrow}} \bar{e}^i\bar{f}\tau fe^jSS \overset{g(\pi_2)g(\pi_3)\cdots g(\pi_m)}{\Longrightarrow} \bar{e}^i\bar{f}\tau fe^j\gamma_1 S \overset{g(\pi_{m+1})g(\pi_{m+2})\cdots g(\pi_n)}{\Longrightarrow} \bar{e}^i\bar{f}\tau fe^j\gamma_1\gamma_2 = \gamma$，$fe^i\gamma = fe^i\bar{e}^i\bar{f}\tau fe^j\gamma_1\gamma_2 \overset{*}{\approx} \tau fe^j\gamma_1\gamma_2 \overset{*}{\approx} \tau S\gamma_2 \approx \gamma_2 \overset{*}{\approx} S$ となる．すると，γ_2 は S から導かれるので P_S, P_R の定義より，$1 < h < l$ であるような h と $\phi \in \Omega^T$ が存在して $\gamma_2 = e^h f\phi$ と表される．このとき，関係 $\sim$ と $\approx$ の定義から γ_2 が S に還元されることはない．ところが，これは $\gamma_2 \overset{*}{\approx} S$ に矛盾する．したがって，仮定は誤り．よって，$g(\pi_1) = S \to \bar{e}^i\bar{f}\tau fe^jSS$ のとき前提は成立しない．また，$g(\pi_1) = S \to \bar{e}^i\bar{f}\tau fe^jSS$ のときと，$g(\pi_1) = S \to \bar{e}^i\bar{f}\theta S fe^j S$ のときの対称性から，(2) のときも前提は成立しない．

以下で，$g(\pi_1) = S \to \bar{e}^i\bar{f}fe^kfe^jS$ の場合を考える．このとき，導出は次のように表される．$S \overset{g(\pi_1)}{\underset{G_R}{\Rightarrow}} \bar{e}^i\bar{f}fe^kfe^jS \overset{g(\pi_2)g(\pi_3)\cdots g(\pi_m)}{\underset{G_R}{\Longrightarrow}} \bar{e}^i\bar{f}fe^kfe^j\delta_1[S] \overset{g(\pi_{m+1})g(\pi_{m+2})\cdots g(\pi_n)}{\underset{G_R}{\Longrightarrow}} \bar{e}^i\bar{f}fe^kfe^j\delta_1[\gamma_1] = \gamma$．ただし，$\delta_1(x) \in \Omega^T$ は，$S \overset{g(\pi_2)g(\pi_3)\cdots g(\pi_m)}{\underset{G_R}{\Longrightarrow}} \delta_1[S]$，$fe^j\delta_1 \overset{*}{\approx} x$ を満たし，$\gamma_1 \in \Omega^T$ は $S \overset{g(\pi_{m+1})g(\pi_{m+2})\cdots g(\pi_n)}{\underset{G_R}{\Longrightarrow}} \gamma_1$，$fe^k\gamma_1 \overset{*}{\approx} S$ を満たすものとする．また，$yield(\delta_1) = u_1xv_1, yield(\gamma_1) = w'$ とおくと，$w = u_1w'v_1$ と表される．

すると，帰納法の仮定より，$C_j x \overset{\pi_2\pi_3\cdots\pi_m}{\underset{G_S}{\Longrightarrow}} \beta_1$，$yield(\beta_1) = uxv$ を満たす $\beta_1(x) \in \Sigma^T$ と $C_k \overset{\pi_{m+1}\pi_{m+2}\cdots\pi_n}{\underset{G_S}{\Longrightarrow}} \alpha_1$，$yield(\alpha_1) = w'$ を満たす $\alpha_1 \in \Sigma^T$ が存在することがいえる．これにより $r(C_k) = 0$ も示されたことに注意する．$r(C_k) = 0$ であることと P_R の構成法より，$\pi_1 = C_i \to C_jC_k \in P_S$．以上より，

$C_i \overset{\pi_i}{\underset{G_S}{\Longrightarrow}} C_jC_k \overset{\pi_2\pi_3\cdots\pi_m}{\underset{G_S}{\Longrightarrow}} \beta_1[C_k] \overset{\pi_{m+1}\pi_{m+2}\cdots\pi_n}{\underset{G_S}{\Longrightarrow}} \beta_1[\alpha_1] = \alpha$ であるような $\alpha \in \Sigma^T$ が存在し，$yield(\alpha) = yield(\beta_1[\alpha_1]) = uyield(\alpha_1)v = uw'v = w$ を満たす．よって成立．

(2) の帰納ステップ　$n > 1$ とする．Σ_0 上の文字列 u,v に対して，$S \overset{g(\pi_1)g(\pi_2)\cdots g(\pi_n)}{\underset{G_R}{\Longrightarrow}} \delta[S]$，$\delta \overset{*}{\approx} x$，$yield(\delta) = uxv$ を満たす $\delta \in \Omega^T$ が存在するとする．この導出は，$g(\pi_1) = S \to e^i f\tau fe^k SS$ の場合，$g(\pi_1) = S \to e^i f\theta S fe^k S$ の場合，$g(\pi_1) = S \to e^i ffe^k fe^j S$ の場合 の3つに場合分けされる．$g(\pi_1) = S \to e^i f\tau fe^k SS$ の場合，導出は，$S \overset{g(\pi_1)}{\underset{G_R}{\Longrightarrow}} \bar{e}^i \bar{f}\tau fe^j SS \overset{g(\pi_2)g(\pi_3)\cdots g(\pi_n)}{\underset{G_R}{\Longrightarrow}} \bar{e}^i \bar{f}\tau fe^j \gamma_1 S = \delta[S]$ と表すことができる．ここで，$\gamma_1 \in \Omega^T$ は，$S \overset{g(\pi_2)g(\pi_3)\cdots g(\pi_n)}{\underset{G_R}{\Longrightarrow}} \gamma_1$，$fe^j\gamma_1 \overset{*}{\approx} S$, $yield(\gamma_1) = u$ を満たすものとする．また，$v = \lambda$ である．すると，帰納法の仮定により，$C_j \overset{\pi_2\pi_3\cdots\pi_n}{\underset{G_S}{\Longrightarrow}} \alpha_1$，$fe^j\alpha_1 \overset{*}{\approx} S$，$yield(\alpha_1) = u$ を満たす $\alpha_1 \in \Sigma^T$ が存在する．また，P_R の構成法より，$\pi_1 = C_ix \to \tau C_jx \in P_S$ である．以上より，$C_ix \overset{\pi_1}{\underset{G_S}{\Longrightarrow}} \tau C_jx \overset{\pi_2\pi_3\cdots\pi_n}{\underset{G_S}{\Longrightarrow}} \tau\alpha_1 x = \beta$ を満たすような $\beta \in \Sigma^T$ が存在し，$yield(\beta) = yield(\tau\alpha_1 x) = yield(\alpha_1)yield(x) = ux = uxv$ となる．よって，$g(\pi_1) = S \to e^i f\tau fe^k SS$ のとき成立．また，$g(\pi_1) = S \to e^i f\tau fe^k SS$ の場合と $g(\pi_1) = S \to e^i f\theta S fe^k S$ の場合には対称性があるので，直ちに $g(\pi_1) = S \to e^i f\theta S fe^k S$ の場合も成立することがいえる．

続いて，$g(\pi_1) = S \to e^i fe^k e^j S$ の場合を示す．このとき導出は，$S \overset{g(\pi_1)}{\underset{G_R}{\Longrightarrow}} \bar{e}^i \bar{f} fe^k fe^j S \overset{g(\pi_2)g(\pi_3)f\cdots g(\pi_m)}{\underset{G_R}{\Longrightarrow}} \bar{e}^i \bar{f} fe^k fe^j \delta_1[S] \overset{g(\pi_{m+1})g(\pi_{m+2})\cdots g(\pi_n)}{\underset{G_R}{\Longrightarrow}} \bar{e}^i \bar{f} fe^k fe^j \delta_1[\delta_2[S]] = \delta[S]$ と表される．ここで，$1 < m < n$ とし，$\delta_1 \in \Omega^T$ は，$S \overset{g(\pi_2)g(\pi_3)\cdots g(\pi_m)}{\underset{G_R}{\Longrightarrow}} \delta_1[S]$，$fe^j\delta_1 \overset{*}{\approx} x$ を満たし，$\delta_2 \in \Omega^T$ は，$S \overset{g(\pi_{m+1})g(\pi_{m+2})\cdots g(\pi_n)}{\underset{G_R}{\Longrightarrow}} \delta_2[S]$，$fe^k\delta_2 \overset{*}{\approx} x$ を満たすものとする．また，$yield(\delta_1) = u_1xv_1$，$yield(\delta_2) = u_2xv_2$ とおくと，$yield(\delta) = u_1u_2xv_2v_1 = uxv$ と表される．すると，帰納法の仮定より，$C_jx \overset{\pi_2\pi_3\cdots\pi_m}{\underset{G_S}{\Longrightarrow}} \beta_1$，$yaald(\beta_1) = u_1xv_1$，を満たす $\beta_1(x) \in \Sigma^T$ と，$C_kx \overset{\pi_{m+1}\pi_{m+2}\cdots\pi_n}{\underset{G_S}{\Longrightarrow}} \beta_2$，$yaald(\beta_2) = u_2xv_2$ を満たす $\beta_2(x) \in \Sigma^T$ が存在する．ここで $r(C_k) = 1$ であることも示されたことに注意する．$r(C_k) = 1$ であることと P_R の構成法より，$\pi_1 = C_ix \to C_jC_kx \in P_S$．以上より，$C_ix \overset{\pi_1}{\underset{G_S}{\Longrightarrow}} C_jC_kx \overset{\pi_2\pi_3\cdots\pi_m}{\underset{G_S}{\Longrightarrow}} \beta_1[C_kx] \overset{\pi_{m+1}\pi_{m+2}\cdots\pi_n}{\underset{G_S}{\Longrightarrow}} \beta_1[\beta_2] = \beta$ で導出され，$yield(\beta) = yield(\beta_1[\beta_2]) = u_1yield(\beta_2)v_1 = u_1u_2xv_2v_1 = uxv$ を満たすような $\beta \in \Sigma^T$ が存在する．よって，$g(\pi_1) = S \to e^i fe^k e^j S$ のときも成立．

ゆえに十分条件のときも示されたので定理が成り立つことが証明された．

5 むすび

Spine 木言語のクラスが，認識可能集合のクラスとの共通集合において閉じていることは形式言語理論によくある方法によって示すことができる．したがって，任意の R に対する $yield(\mathrm{R} \cap \mathrm{D})$ が TAL であるということがいえる．このことと定理 4.2 により，TALs は，認識可能集合 D と拡張 Dyck 言語 R によって特徴付けられた言語のクラスであるといえる．

参考文献

[1] W.S.Brainerd,"Tree generating regular systems" Inf.Control,vol.14,no.2,pp.217-131,1969.

[2] N.Chomsky,"Context-free grammars and pushdown storage",Quarterly Progress Report No.65,pp.187-194.Massachusetts Institute of Technology Research Laboratory of Electronics,April 1962.

[3] A.Fujiyoshi,T.Kasai. "Spinal-formed context-free tree grammars",Theory of Computing Systems vol33,pp.59-83,2000.

[4] T.Kasai,"A universal context-free grammar" Inf.Control,28(1):30-34, May 1975.

[5] A.K,Joshi,L.S.Levy,and M.Takahashi"Tree adjunct grammars",J.Comput.Syst.Sci.,vol.10,no.1,pp.136-163,1975.

[6] K.Vijay-Shanker,D.J.Weir,"The equivalence of four extenseions of context-free grammars", Mathematical Systems Theory,vol.27,no.6,pp.511-546,1994.

[7] J.Engelfriet,E.M.Schmidt"IO and OI" Journal of Computer and System Sciences,vol.15,no.3,pp.328-353,1977,and vol.16,no.1,pp.67-99,1978.