数理解析研究所講究録
1093 巻 1999 年 218-223



# 一意並列解析可能ユニフィケーション文法

## A Uniquely Parallel Parsable Unification Grammar

李 佳, 森田 憲一, 岩本 宙造, 今井 克暢
広島大学工学部
{lijia, morita, iwamoto, imai}@ke.sys.hiroshima-u.ac.jp

## 1 はじめに

一意解析可能文法 (Uniquely Parsable Grammar, UPG) は、書換規則にある種の明示的な条件をつけることにより、構文解析をバックトラックなしに (一意に) 進めるようにした文法として、森田ら [1] によって提案された。この制約条件にもかかわらず UPG は Turing 機械と同等な生成能力を持ち、従って万能である。

本稿では、ユニフィケーション文法 (UG) の考えを UPG に取り入れて拡張した一意解析可能ユニフィケーション文法に注目し、その一意解析可能性に加え、構文解析を効率的に行える並列解析が実現できるように、各書換規則に文脈インデックスを導入し、一意並列解析可能ユニフィケーション文法 (Uniquely Parallel Parsable Unification Grammar, UPPUG) を定義する。またその (直列的) 一意解析可能性に加え、並列解析可能性も合わせて示す。

## 2 一意解析可能文法

一意解析可能文法 UPG は、構文解析をバックトラックなしに行うことができる句構造文法である。ここで構文解析 (parsing) とは、文法 $G$ に対して、与えられた記号列が $G$ によって生成される言語に属するか否かの判定を行うと共に、生成される場合には導出過程を求めることを言う。しかし、書換規則を逆適用し、記号列を還元することによって解析を行う際、一般に各ステップにおいて、逆適用可能な規則が複数存在する。そのため、ある規則を選んで解析に失敗した時、別の選択肢があるところまで戻り、別の規則を使う必要がある。これをバックトラックと呼ぶ。構文解析をバックトラックなしに決定的 (一意) に進めることができれば効率的である。

UPG では、書換規則に次のような制約条件 (UPG 条件) を設けている: まず、ある規則の右辺の接頭辞と、別のある規則の右辺の接尾辞とが一致するならば、それらの共通部分が、各規則の左辺において、それぞれ同じ位置になくてはならない。さらに、任意の規則の右辺が別の規則の右辺に含まれることを禁ずる。このような制約から、逆適用可能な規則が複数あった場合、それらがそれぞれの逆適用位置において重なる部分は、いずれの規則を逆適用しても書き換えられることがない。従って、UPG の還元は一種の合流性を満たし、バックトラックを行わずに一意に進むことができる。

さらに、森田らは、決定性 Turing 機械 (DTM) における言語受理のプロセスを、UPG 条件を満たす書換規則による語の還元過程によって、完全に模倣できることを示した。すなわち、UPG の言語クラスと、DTM によって受理される言語のクラスとが同等である。また、決定性 Chomsky 階層によって特徴づけられる UPG の三つのサブクラス [1] も定義した。

## 3 ユニフィケーション文法

ユニフィケーション文法 UG は生成文法の各非終端記号に引数を持たした文法である。これは引数間でユニフィケーションを行うことによって、生成還元を実行する文法である。また、引数を利用して、情報の伝達や保持が簡潔に行える特徴を持つ。これによって、ユニフィケーション文法は文法の記述力を高め、構文の曖昧性をうまく解消することができる。

特に、構文解析の効率の良さおよび文法構成の便利さから、自然言語処理の分野 [5] において注目されてきた文脈自由文法に対し、自然言語における互いに割り込んだ構文要素間の依存関係 (例えば、英語の respectively 構文) を適切に表現できないと指摘された。また、自然言語の統語構造を文脈自由言語で記述しようとすると、類似の規則が非常に多数必要となる。これらの問題の多くはユニフィケーション文法によって解決できる。すなわち、文脈自由文法の非終端記号に素性構造 (引数) を持たせ、比較的少数の規則型によって、統語構造を簡潔に記述することができる。

### 3.1 項およびユニフィケーション

**定義 3.1** ユニフィケーション文法 UG において使われる記号は、関数記号と変数の 2 種類である。関数記号 $f$ にはその引数の個数を示す非負整数 (アリティ)$n$ が付随して $f^{(n)}$ とも表される。アリティ0の関数記号は定数と呼ばれる。$F$ と $V$ をそれぞれ関数記号、変数の集合とし、また $F$ 中の定数の集合を $Const(F)$ と表す。$F$ と $V$ 上の項は次のように再帰的に定義される

(1) 変数は項である
(2) 定数は項である
(3) $f^{(n)}$ $(\in F)$ が関数記号で、$t_1, \cdots, t_n$ が項ならば、$f^{(n)}(t_1, \cdots, t_n)$ も項である

項全体の集合を $Term(F, V)$ と表し、また変数以外の項を複合項と呼んでその集合を $CTerm(F, V)$ と記す。すなわち、$CTerm(F, V) = Term(F, V) - V$。 □

**定義 3.2** 有限個の複合項を一列に並べたものを、複合項

列と呼び、その集合を $CTerm(F,V)^*$ と表す。さらに、

$$CTerm(F,V)^+ = CTerm(F,V)^* - \{\varepsilon\}$$

とする ($\varepsilon$ は空列)。$\eta = s_1, \cdots, s_n \in CTerm(F,V)^+$ に対して、

$$head(\eta) = s_1$$
$$tail(\eta) = s_n$$

とする (但し、$head(\varepsilon) = tail(\varepsilon) = \varepsilon$)。また、$Var(t)$ は、複合項列 $t \in CTerm(F,V)^*$ に含まれる変数の集合を表す。 □

**定義 3.3[代入]** $F$, $V$ をそれぞれ関数記号および変数の集合とする時、写像 $\sigma: V \to Term(F,V)$ を代入 (substitution) と呼ぶ。また便宜上、代入を以下の集合によって表す。但し、$X_i \in V$, $t_i \in Term(F,V)$。

$$\{t_i/X_i \mid \sigma(X_i) = t_i \wedge X_i \neq t_i\}$$

そして、

$$Dom(\sigma) = \{X_i \mid \sigma(X_i) \neq X_i\}$$

とする。項 $s$ に対する代入 $\sigma$ の適用を以下のように定義し、その結果を $s\sigma$ と表す。但し、$f^{(n)} \in F$、かつ $X_i \in V$, $r_i \in Term(F,V)$ $(1 \leq i \leq n)$。

(1) $s = X_i$ ならば、$s\sigma = \sigma(X_i)$

(2) $s = f^{(n)}(r_1, \cdots, r_n)$ ならば、$s\sigma = f^{(n)}(r_1\sigma, \cdots, r_n\sigma)$

複合項列 $s_1, \cdots, s_m$ への代入 $\sigma$ の適用を次のように定義する

$$(s_1, \cdots, s_m)\sigma = s_1\sigma, \cdots, s_m\sigma$$

また、二つの代入 $\sigma$ と $\tau$ の合成 (composition) $\sigma \circ \tau$ は次のようなものである。

$$s(\sigma \circ \tau) = (s\sigma)\tau$$

最後に、変数の改名 (renaming of variables) と呼ばれる特別な代入は単射 $\sigma: V \to V$ である。 □

**定義 3.4[ユニフィケーション]** 項 $s$, $t \in Term(F,V)$ がユニフィケーション可能 (unifiable) とは、$s\sigma = t\sigma$ となる代入 $\sigma$ が存在することを言い、$s \sim t$ で表す。同様に、複合項列 $\alpha$, $\beta \in (CTerm(F,V)^*$ に対して、ある代入 $\sigma$ により $\alpha\sigma = \beta\sigma$ となるならば、$\alpha$ と $\beta$ がユニフィケーション可能と言い、$\alpha \sim \beta$ と表す。このような $\sigma$ をユニフィケーション作用素 (unifier) と呼ぶ。 □

## 3.2 Unification Grammar の定義

**定義 3.5 [UG]** ユニフィケーション文法 UG は

$$G = (F, T, V, P, s^{(n_s)})$$

によって定義される。但し、記号 $F$, $T$, $V$, $P$, $s^{(n_s)}$ は次の通りである。

(1) $F$ は関数記号の集合で、$Const(F) \neq \emptyset$

(2) $T (\subset Const(F))$ は終端記号の集合で、$T \neq \emptyset$

(3) $V$ は可算個の変数の集合

(4) $s^{(n_s)} (\in F)$ は開始記号

そして、$P$ は複合項列を書換えるための、規則の有限集合である。各規則 $R (\in P)$ は次の形で与えられる。

$$R = \alpha \to \beta$$

但し、$\alpha \in CTerm(F,V)^+ - T^+$, $\beta \in CTerm(F,V)^*$ かつ $Var(\alpha) \subseteq Var(\beta)$。

便宜上、規則に現れる変数の集合を $Var(R)$ とする (従って、$Var(R) = Var(\beta)$)。同様に、規則 $R$ への代入 $\sigma$ の適用を $R\sigma (= \alpha\sigma \to \beta\sigma)$ と書く。 □

**定義 3.6[導出]** UG $G = (F, T, V, P, s^{(n_s)})$ において、複合項列 $\xi = r_1, \cdots, r_k \in CTerm(F,V)^*$ に対して、規則 $R = s_1, \cdots, s_m \to t_1, \cdots, t_n \ (\in P)$ が $\xi$ の位置 $j$ で適用可能とは

$$(r_j, \cdots, r_{j+m-1})\sigma = (s_1, \cdots, s_m)\sigma$$

となる代入 $\sigma$ が存在することを言う。但し、ここで $Var(\xi) \cap Var(R) = \emptyset$ と仮定する (そうでない時、$R$ に対して変数の改名を行う)。また、

$$\eta = (r_1, \cdots, r_{j-1}, t_1, \cdots, t_n, r_{j+m}, r_k)\sigma$$

であるならば、$G$ において $\xi$ から $\eta$ が直接導出されると言い、$\xi \underset{G}{\Rightarrow} \eta \ [R, j, \sigma]$、あるいは $\xi \overset{[R,j,\sigma]}{\underset{G}{\Rightarrow}} \eta$ と表記する ($G$ は省略可能)。また、$[R, j, \sigma]$ を書換えのラベルと呼ぶ。

なお、導出の反射的推移的閉包を $\overset{*}{\Rightarrow}$ で、$n$ ステップの導出を $\overset{n}{\Rightarrow}$ で表す。$X_i \in V$ $(1 \leq i \leq n_s)$ を相異なる $n_s$ 個の変数とする時、

$$s(X_1, \cdots, X_{n_s}) \overset{*}{\underset{G}{\Rightarrow}} \eta$$

ならば、複合項列 $\eta$ は $G$ から導出 (生成) 可能と言い、$G$ の文形式と呼ぶ。文形式 $\eta$ が $G$ の文となるための必要かつ十分条件は $\eta \in T^*$ である。そして、$G$ によって生成される言語 $L(G)$ はすべての文の集合である。

$$L(G) = \{w \mid s(X_1, \cdots, X_{n_s}) \overset{*}{\underset{G}{\Rightarrow}} w \wedge w \in T^*\}$$ □

**定義 3.7[還元]** UG $G = (F, T, V, P, s^{(n_s)})$ において、複合項列 $\xi = r_1, \cdots, r_k \in CTerm(F,V)^*$ に対して、規則 $R = s_1, \cdots, s_m \to t_1, \cdots, t_n \ (\in P)$ が $\xi$ の位置 $j$ で逆適用可能と言われるのは

$$r_j, \cdots, r_{j+n-1} = (t_1, \cdots, t_n)\sigma$$

となる代入 $\sigma$ が存在することを言う。また、

$$\eta = r_1, \cdots, r_{j-1}, (s_1, \cdots, s_m)\sigma, r_{j+n}, \cdots, r_k$$

であるならば、$G$ において $\eta$ が $\xi$ から直接還元されると言い、$\xi \underset{G}{\Leftarrow} \eta \ [R, j, \sigma]$、あるいは $\xi \overset{[R,j,\sigma]}{\underset{G}{\Leftarrow}} \eta$ と表記する ($G$ は省略可能)。導出と同様に $[R, j, \sigma]$ を書換えのラベルと呼ぶ。

なお、還元の反射的推移的閉包を、また $n$ ステップの還元をそれぞれ $\overset{*}{\Leftarrow}$, $\overset{n}{\Leftarrow}$ と表す。ある項 $u_1, \cdots, u_{n_s} \in Term(F,V)$ に対し

$$\xi \overset{*}{\underset{G}{\Leftarrow}} s(u_1, \cdots, u_{n_s})$$

ならば、$\xi$ が $G$ において開始記号まで還元できると言う。 □

**定義 3.8** UG $G = (F, T, V, V, P, s^{(n_s)})$ において、直接還元 $\xi \Leftarrow \eta[R, i, \sigma]$ が直接最左還元と呼ばれるのは、ある複合項列 $\eta' \in CTerm(F,V)^*$ により、

$$\xi \Leftarrow \eta'[R', i', \sigma']$$

となる任意の $[R', i', \sigma']$ に対して、$i \leq i'$ が成り立つことを言う。ここで、直接最左還元を

$$\xi \underset{lmr}{\Leftarrow} \eta[R, i, \sigma]$$

と表記する。また、還元 $\eta \Leftarrow \xi_1 \Leftarrow \cdots \Leftarrow \xi_n$ が最左還元であるとは、

$$\eta \underset{lmr}{\Leftarrow} \xi_1 \underset{lmr}{\Leftarrow} \cdots \underset{lmr}{\Leftarrow} \xi_n$$

となることを言う。同様に、最左還元の反射的推移的閉包、および $n$ ステップの最左還元をそれぞれ $\overset{*}{\underset{lmr}{\Leftarrow}}$, $\overset{n}{\underset{lmr}{\Leftarrow}}$ と表す。 □

複合項列へ規則を適用する導出、あるいは逆適用による還元のいずれも、まずユニフィケーション操作を行った後、書換えを実行するものであることから、UGにおける導出と還元の関係は一般の句構造文法ほど自明でない。しかしながら、任意の複合項列が文法$G$によって生成されることと、その複合項列が$G$において開始記号まで還元できることとが等価であることが示せる。

**補題 3.1** $G=(F,T,V,P,s^{(n_s)})$をUGとする。また、$\xi \in (CTerm(F,V)^*$を任意の複合項列、$t_1,\cdots,t_{n_s} \in Term(F,V)$を項とする。もし、

$$\xi \overset{*}{\Leftarrow} s(t_1,\cdots,t_{n_s})$$

ならば、任意の代入$\sigma$に対して、

$$\xi\sigma \overset{*}{\Leftarrow} s(t_1\sigma,\cdots,t_{n_s}\sigma)$$

が成り立つ。

**(証明)** 還元ステップ$n$に関する帰納法を用いる。まず$n=0$の場合、補題は自明である。$n=k-1$の時に補題の成立を仮定し、次の$k$ステップの還元を考える。

$$\xi \overset{[R,j,\tau]}{\Leftarrow} \eta \overset{k-1}{\Leftarrow} s(t_1,\cdots,t_{n_s})$$

規則$R=\alpha\rightarrow\beta$と仮定すると、ある$\gamma,\lambda \in CTerm(F,V)^*$に対して、$\xi=\gamma,\beta\tau,\lambda$と書ける。従って、

$$\xi=\gamma,\beta\tau,\lambda \overset{[R,j,\tau]}{\Leftarrow} \gamma,\alpha\tau,\lambda=\eta$$

となり、それゆえ、任意の代入$\sigma$に対して、

$$\xi\sigma=\gamma\sigma,\beta(\tau\circ\sigma),\lambda\sigma \overset{[R,j,\tau\circ\sigma]}{\Leftarrow} \gamma\sigma,\alpha(\tau\circ\sigma),\lambda\sigma=\eta\sigma$$

となる。帰納法の仮定により、$\eta\sigma \overset{*}{\Leftarrow} s(t_1\sigma,\cdots,t_{n_s}\sigma)$であるので、

$$\eta\sigma \overset{*}{\Leftarrow} s(t_1\sigma,\cdots,t_{n_s}\sigma)$$

が成り立つ。 □

**補題 3.2** $G=(F,T,V,P,s^{(n_s)})$をUGとする。ここで、$\xi \in CTerm(F,V)^*$を複合項列、$X_1,\cdots,X_{n_s} \in V$を相異なる変数とする時、

$$s(X_1,\cdots,X_{n_s}) \overset{*}{\Rightarrow} \xi$$

ならば、項$t_1,\cdots,t_{n_s} \in Term(F,V)$が存在し、

$$\xi \overset{*}{\Leftarrow} s(t_1,\cdots,t_{n_s})$$

が成り立つ。

**(証明)** 導出ステップ数$n$に関する帰納法を用いる。まず$n=0$ならば、補題は自明である。$n=k-1$において補題の成立を仮定し、次の$k$ステップの導出を考える。

$$s(X_1,\cdots,X_{n_s}) \overset{k-1}{\Rightarrow} \eta \overset{[R,j,\tau]}{\Rightarrow} \xi$$

ある$\gamma,\alpha',\lambda \in CTerm(F,V)^*$に対して、$\eta=\gamma,\alpha',\lambda$とする。また、$R=\alpha\rightarrow\beta$、$Var(\eta)\cap Var(R)=\emptyset$と仮定する(そうでない時、$R$に対する変数の改名を行う)。導出の定義から、$\alpha'\tau=\alpha\tau$、しかも$\xi=\gamma\tau,\beta\tau,\lambda\tau$となる。従って、$\xi$は$R$によって還元可能であり、

$$\xi=\gamma\tau,\beta\tau,\lambda \overset{[R,j,\tau]}{\Leftarrow} \gamma\tau,\alpha\tau,\lambda\tau=\eta\tau$$

が成り立つ($\alpha'\tau=\alpha\tau$であるので)。従って、補題3.1および帰納法の仮定により、

$$\xi \overset{*}{\Leftarrow} s(t_1,\cdots,t_{n_s})$$

となる項$t_1,\cdots,t_{n_s}$が存在する。 □

**補題 3.3** $G=(F,T,V,P,s^{(n_s)})$をUGとする。もし、任意の複合項列$\xi \in CTerm(F,V)^*$に対して、ある項$t_1,\cdots,t_{n_s} \in Term(F,V)$が存在し、

$$\xi \overset{*}{\Leftarrow} s(t_1,\cdots,t_{n_s})$$

ならば、相異なる変数$X_1,\cdots,X_{n_s} \in V$に対して、

$$s(X_1,\cdots,X_{n_s}) \overset{*}{\Rightarrow} \xi$$

が成り立つ。

**(証明)** 還元ステップ数$n$に関する帰納法を用いる。まず$n=0$ならば、補題は自明である。$n=k-1$の時に、補題の成立を仮定し、次の$k$ステップの還元を考える。

$$\xi \overset{[R,j,\tau]}{\Leftarrow} \eta \overset{k-1}{\Leftarrow} s(t_1,\cdots,t_{n_s})$$

$R=\alpha\rightarrow\beta$とし、$Var(\xi)\cap Var(R)=\emptyset$と仮定する(そうでない時、任意の変数の改名$\sigma$に対して、$\xi \Leftarrow \eta\,[R\sigma,j,\sigma^{-1}\circ\tau]$が成り立つので、$Var(R\sigma)\cap Var(\xi)=\emptyset$となる$R\sigma$を$\xi$の位置$j$で逆適用する。但し、変数の改名は単射で、かつ$Var(R)=Dom(\sigma)$と仮定しても一般性を失わない)。

還元の定義から、ある$\gamma,\lambda \in CTerm(F,V)^*$が存在し、

$$\xi=\gamma,\beta\tau,\lambda \overset{[R,j,\tau]}{\Leftarrow} \gamma,\alpha\tau,\lambda=\eta$$

と書ける。一方、$Var(\xi)\cap Var(R)=\emptyset$なので、$Var(\xi)\cap Dom(\tau)=\emptyset$としても一般性を失わない。従って、$\eta\tau=\eta$となり、それゆえ、$\eta$には$R$が位置$j$で適用可能で、

$$\eta=\gamma,\alpha\tau,\lambda \overset{[R,j,\tau]}{\Rightarrow} \gamma,\beta\tau,\lambda$$

が成り立つ。帰納法の仮定により、$s(X_1,\cdots,X_{n_s}) \overset{*}{\Rightarrow} \eta$であるので、

$$s(X_1,\cdots,X_{n_s}) \overset{*}{\Rightarrow} \xi$$

が成り立つ。 □

以上の補題から、文法$G$から生成されうる複合項列の集合と、同じ$G$において開始記号まで還元できるものの集合とが一致する。従って、書換規則を逆適用し、還元を行うことで、$G$の文を含む任意の複合項列の生成可能性を解析することができる。

## 3.3 文脈インデックスの導入

本研究では、UPPUGにおける書換規則の制約を明示的に記述するため、またその並列的な解析を定義するために、文脈インデックス付き書換規則を次のように導入する。

**定義 3.9** UG $G=(F,T,V,P,s^{(n_s)})$において、$l,r$を非負の整数とする。規則$\alpha\rightarrow\beta$に対し、$\alpha_1,\beta_1,\gamma,\delta \in CTerm(F,V)^*$が存在し、$\alpha=\gamma\alpha_1\delta$, $\beta=\gamma\beta_1\delta$となり、かつ$\alpha_1 \neq \varepsilon$, $\beta_1 \neq \varepsilon$の時、$head(\alpha_1) \neq head(\beta_1)$, $tail(\alpha_1) \neq tail(\beta_1)$ならば、$\gamma,\delta$をそれぞれ規則の左文脈、右文脈と呼ぶ。さらに、$|\gamma|=l, |\delta|=r$とする時、$(l,r)$を文脈インデックス、

$$[\alpha\rightarrow\beta,\ (l,r)]$$

を文脈インデックス付き書換規則と呼ぶ。 □

UPGと並んで、森田らは一意解析可能アレイ文法UPAG(Uniquely Parsable Array Grammar) [2]を提案した。UPAGでは、書換規則にその規則の左辺および右辺が同形であるという制約を加え、左辺と右辺で書き換わらない部分を文脈部(context portion)と呼ぶ。さらに、UPGと同様に、ある二つの規則の右辺を重ねあわせた時に、一致する部分は各規則の左辺においても、同じ場所になくてはならない、従って、規則の文脈の一部分である。これらの制約により、UPAGに対し、規則の逆適用を同時に行う並列還元を容易に定義できる。

しかし、UPGでは、ある規則の右辺の接頭辞が他(または同一)の規則の右辺の接尾辞と同じならば、それらの部分は規則それぞれの左辺の接頭辞、接尾辞となるが、文脈の一部分にあたると限らない。それは、一次元生成

文法では、規則の左右の長さが一般に違うので、文脈の指定は必ずしも一意に決められない。例えば、$A \to ABA$ を考えると、$A$ は左文脈、あるいは右文脈のどちらとも呼べる。$AB \to ACAB$ についても同様である。従って、UPG では並列還元を行う操作が定義されていない。

そこで、一意並列解析可能ユニフィケーション文法 UPPUG においては、各書換規則に文脈インデックスを付随させ、規則の文脈の部分を明確に指定する。さらに、UPG における制約と同様の制約をインデックス付き規則に対して課す。これによって、同時に逆適用可能な二つの規則が逆適用位置において重なった部分は、常に両方の規則の文脈にあり、UPAG と同様に、各規則を並列に逆適用することができる。

# 4 UPPUG

**定義 4.1[UPPUG]** 一意並列解析可能ユニフィケーション文法 UPPUG は $G = (F, T, V, P, s^{(n_s)}, \$)$ で定義される。$F, T, V, s^{(n_s)}$ の定義は UG と同様で、\$は特別な境界記号 ($\$ \notin F \cup V$) を表す。$P$ は文脈インデックス付き書換規則の有限集合で、$\alpha \in CTerm(F,V)^+ - T^+$, $\beta \in CTerm(F,V)^+$ に対し、各規則は次のいずれかの形で与えられる。

$$[\alpha \to \beta,\ (l,r)],\ [\$\alpha \to \$\beta, (l,r)]$$
$$[\alpha\$ \to \beta\$, (l,r)],\ [\$\alpha\$ \to \$\beta\$, (l,r)]$$

$P$ はさらに以下の条件を満たす。

**UPPUG 条件:**

(1) $P$ に属する各書換規則の右辺 $\beta$ は $s(X_1, \cdots, X_{n_s})$, $\$s(X_1, \cdots, X_{n_s})$, $s(X_1, \cdots, X_{n_s})\$$, あるいは $\$s(X_1, \cdots, X_{n_s})\$$ のいずれとも単一化可能でない。

(2) $P$ 中の任意の書換規則対 $R_1 = [\alpha_1 \to \beta_1,\ (l_1, r_1)]$, $R_2 = [\alpha_2 \to \beta_2,\ (l_2, r_2)]$ に対して ($R_1 = R_2$ の時も同様)、以下が成立する。

(a) ある $\delta_1, \delta_2, \beta_1', \beta_2', \in (CTerm(F,V) \cup \{\$\})^+$ に対し、もし $\beta_1 = \delta_1\beta_1'$, $\beta_2 = \beta_2'\delta_2$、しかも $\delta_1 \sim \delta_2$ ならば、$l_1 \ge |\delta_1|$, $r_2 \ge |\delta_2|$

(b) ある $\gamma, \delta \in (CTerm(F,V) \cup \{\$\})^*$ に対して, もし $\beta_1 \sim \gamma\beta_2\delta$ ならば、$R_1 = R_2$ □

UPPUG-条件 2(*a*) は、ある規則の右辺にある接頭辞が、別 (または同一) のある規則の右辺の接尾辞とユニフィケーション可能ならば、規則の左辺もそれらの部分を、それぞれ接頭辞および接尾辞として持たなくてはならないと述べている。しかも、それらの部分はいずれの規則においても、文脈インデックスが示す文脈部になくてはならない。そして、条件 2(*b*) によって、どの規則の右辺も、他の規則の右辺にある部分複合項列とユニフィケーション可能でない。

次に、言語の生成および解析に関わる UPPUG の導出や還元を UG 同様に定義することができる。但し、UPPUG において複合項列は、常に両端が\$によって囲まれ、$\$\eta\$$ ($\eta \in CTerm(F,V)^*$) の形で扱われる。また、複合項列 $\xi \in CTerm(F,V)^+$ に対して、相異なる変数 $X_1, \cdots, X_{n_s} \in V$ が存在し、

$$\$s(X_1, \cdots, X_{n_s})\$ \overset{*}{\underset{G}{\Rightarrow}} \$\xi\$$$

ならば、$\xi$ は $G$ から生成可能と言う。また、ある項 $t_1, \cdots, t_{n_s} \in Term(F,V)$ に対して、

$$\$\xi\$ \overset{*}{\underset{G}{\Leftarrow}} \$s(t_1, \cdots, t_{n_s})\$$$

が成り立つ時、$G$ において、$\xi$ は開始記号まで還元できると言う。そして、UPPUG の言語 $L(G)$ は UG と同様に、すべての文の集合である。

$$L(G) = \{w \mid \$s(X_1, \cdots, X_{n_s})\$ \overset{*}{\Rightarrow} \$w\$ \wedge w \in T^*\}$$

なお、UPPUG の生成能力が万能、すなわち、Turing 機械により特徴付けられることは、次の定理より導くことができる。

**定理 4.1** UPPUG の族は UPG と同様に、万能な生成文法の族である。

**(証明)** 決定性 Tuing 機械 DTM における語の受理の動作を、UPG の制約を満たす書換規則 (文献 [1] 参照) の逆適用によって、完全に模倣できる。一方、UPPUG は、UPG の非終端記号や、終端記号などを定数として扱うことにより、それらの書換規則を模倣できる。これにより、UPPUG も Turing 機械を模倣することができ、従って万能である。 □

# 5 UPPUG の一意解析性

**補題 5.1[還元の合流性]** $G = (F, T, V, P, s^{(n_s)}, \$)$ を UPPUG、$\eta \in (CTerm(F,V) \cup \{\$\})^+$ を任意の複合項列とする。任意の二つの規則 $R_1 = [\alpha_1 \to \beta_1,\ (l_1, r_1)]$, $R_2 = [\alpha_2 \to \beta_2,\ (l_2, r_2)]$ と、$1 \le i_1 < i_2 \le |\eta|$ を満たす任意の自然数 $i_1, i_2$ に対し、もし、

$$\eta \Leftarrow \xi_1[R_1, i_1, \sigma_1]$$
$$\eta \Leftarrow \xi_2[R_2, i_2, \sigma_2]$$

であるならば、次の条件を満たす複合項列 $\xi \in (CTerm(F,V) \cup \{\$\})^+$ が唯一存在する (但し、$i_2' = i_2 + |\alpha_1| - |\beta_1|$)。

$$\eta \overset{[R_1, i_1, \sigma_1]}{\Leftarrow} \xi_1 \overset{[R_2, i_2', \sigma_2]}{\Leftarrow} \xi$$
$$\eta \overset{[R_2, i_2, \sigma_2]}{\Leftarrow} \xi_2 \overset{[R_1, i_1, \sigma_1]}{\Leftarrow} \xi$$

**(証明)** (1) $i_2 - i_1 \ge |\beta_1|$ の場合

ある $\gamma, \theta, \lambda \in (CTerm(F,V) \cup \{\$\})^*$ に対し、

$$\eta = \gamma, \beta_1\sigma_1, \theta, \beta_2\sigma_2, \lambda$$

と書ける (但し、$|\gamma| = i_1 - 1$, $|\gamma, \beta_1\sigma_1, \theta| = i_2 - 1$)。従って、

$$\xi_1 = \gamma, \alpha_1\sigma_1, \theta, \beta_2\sigma_2, \lambda$$
$$\xi_2 = \gamma, \beta_1\sigma_1, \theta, \alpha_2\sigma_2, \lambda$$

となる。従って、$\xi_1$ と $\xi_2$ には、それぞれ $R_2$, $R_1$ が位置 $i_2'$, $i_1$ において逆適用可能となる。つまり、

$$\xi_1 \overset{[R_2, i_2', \sigma_2]}{\Leftarrow} \gamma, \alpha_1\sigma_1, \theta, \alpha_2\sigma_2, \lambda$$
$$\xi_2 \overset{[R_1, i_1, \sigma_1]}{\Leftarrow} \gamma, \alpha_1\sigma_1, \theta, \alpha_2\sigma_2, \lambda$$

となり、補題が成り立つ ($\xi = \gamma, \alpha_1\sigma_1, \theta, \alpha_2\sigma_2, \lambda$)。

(2) $i_2 - i_1 \le |\beta_1|$ の場合

UPPUG の条件により、ある $\gamma, \lambda, \alpha_1', \beta_1', \alpha_2', \beta_2', \delta', \delta_1, \delta_2 \in (CTerm(F,V) \cup \{\$\})^*$ が存在して、$\eta, R_1, R_2$ は次のように書ける。

$$\eta = \gamma, \beta_1'\sigma_1, \delta', \beta_2'\sigma_2, \lambda$$
$$R_1 = [\alpha_1'\delta_1 \to \beta_1'\delta_1,\ (l_1, r_1)]$$
$$R_2 = [\delta_2\alpha_2' \to \delta_2\beta_2',\ (l_2, r_2)]$$

但し、$\delta' = \delta_1\sigma_1 = \delta_2\sigma_2$, $|\gamma| = i_1 - 1$, $|\gamma, \beta_1'\sigma_1| = i_2 - 1$, $|\delta'| \le min\{r_1, l_2\}$。$\delta_1\sigma_1 = \delta_2\sigma_2$ および $Var(R_1) \cap$

$Var(R_2) = \emptyset$により (そうでない時、$R_1, R_2$ に対して変数の改名を行う)、$\delta_1 \sim \delta_2$ が成り立つ。
従って、

$$\xi_1 = \gamma, \alpha'_1\sigma_1, \delta_1\sigma_1, \beta'_2\sigma_2, \lambda$$
$$\xi_2 = \gamma, \beta'_1\sigma_1, \delta_2\sigma_2, \alpha'_2\sigma_2, \lambda$$

となる。$\delta_1\sigma_1 = \delta_2\sigma_2$ であるので、$\xi_1$, $\xi_2$ にはそれぞれ $R_2$, $R_1$ が位置 $i'_2$, $i_1$ で逆適用可能となる。つまり、

$$\xi_1 \overset{[R_2, i'_2, \sigma_2]}{\Leftarrow} \gamma, \alpha'_1\sigma_1, \delta_2\sigma_2, \alpha'_2\sigma_2, \lambda$$
$$\xi_2 \overset{[R_1, i_1, \sigma_1]}{\Leftarrow} \gamma, \alpha'_1\sigma_1, \delta_1\sigma_1, \alpha'_2\sigma_2, \lambda$$

となり、補題が成り立つ ($\xi = \gamma, \alpha'_1\sigma_1, \delta', \alpha'_2\sigma_2, \lambda$)。 □

**定理 5.1[一意解析可能性]** $G = (F, T, V, P, s^{(n_s)}, \$)$ を UPPUG、$\eta \in (CTerm(F, V) \cup \{\$\})^+$ を $G$ から生成可能な任意の複合項列とする。もし、ある項 $t_1, \cdots, t_{n_s} \in CTerm(F, V)$ に対し、

$$\eta \overset{n}{\Leftarrow} \$s(t_1, \cdots, t_{n_s})\$$$

が成り立つならば、$\eta \Leftarrow \xi\ [R, j, \sigma]$ となるような任意の $[R, j, \sigma]$ に対して、次の還元が成り立つ。

$$\eta \overset{[R, j, \sigma]}{\Leftarrow} \xi \overset{n-1}{\Leftarrow} \$s(t_1, \cdots, t_{n_s})\$$$

**(証明)** 還元回数 $n$ に関する帰納法を用いる。まず $n = 1$ の場合、定理の成立が自明である。$n = k - 1$ の時に一意解析可能性が成り立つと仮定し、次の $k$ ステップの還元を考える。

$$\eta \overset{[R_1, i_1, \sigma_1]}{\Leftarrow} \xi_1 \overset{k-1}{\Leftarrow} \$s(t_1, \cdots, t_{n_s})\$$$

$[R_1, i_1, \sigma_1] \neq [R_2, i_2, \sigma_2]$ ($i_1 \neq i_2$) であるようなラベル $[R_2, i_2, \sigma_2]$ が存在して、

$$\eta \overset{[R_2, i_2, \sigma_2]}{\Leftarrow} \xi_2$$

となる。補題 5.1 により、ある $\xi'$ と $i'_1$, $i'_2$ が存在し、

$$\eta \overset{[R_1, i_1, \sigma_1]}{\Leftarrow} \xi_1 \overset{[R_2, i'_2, \sigma_2]}{\Leftarrow} \xi'$$
$$\eta \overset{[R_2, i_2, \sigma_2]}{\Leftarrow} \xi_2 \overset{[R_1, i'_1, \sigma_1]}{\Leftarrow} \xi'$$

が成り立つ。帰納法の仮定により、

$$\xi_1 \overset{[R_2, i'_2, \sigma_2]}{\Leftarrow} \xi' \overset{k-2}{\Leftarrow} \$s(t_1, \cdots, t_{n_s})\$$$

であるので、

$$\eta \overset{[R_2, i_2, \sigma_2]}{\Leftarrow} \xi_2 \overset{k-1}{\Leftarrow} \$s(t_1, \cdots, t_{n_s})\$$$

が得られる。 □

このように UPPUG において、開始記号まで還元可能 (同時に生成可能) な複合項列に対し、規則をどのように逆適用しても、その逆適用順番に依存せずに、同じ還元ステップ数で決定的に解析を進めることができる。そして、定義 4.4 から、次の系を容易に導くことができる。すなわち、最左還元によって、一意に解析を行うことが可能である。

**系 5.1** $G = (F, T, V, P, s^{(n_s)}, \$)$ を UPPUG、$\eta \in (CTerm(F, V) \cup \{\$\})^+$ を任意の複合項列とする。ある項 $t_1, \cdots, t_{n_s} \in Term(F, V)$ に対して、

$$\eta \overset{n}{\Leftarrow} \$s(t_1, \cdots, t_{n_s})\$$$

となるならば、

$$\eta \underset{lmr}{\overset{n}{\Leftarrow}} \$s(t_1, \cdots, t_{n_s})\$$$

が成り立つ。 □

# 6 UPPUG の一意並列解析性

複合項列 $\eta$ に対して、逆適用可能な二つの規則 $R_1, R_2$ が存在するならば、UPPUG 条件により、$R_1, R_2$ の逆適用位置が重なっても、その部分はいずれの規則の逆適用によっても書換えられることがない。さらに、還元の定義 4.3 より、$\eta$ に対する書換は、規則ごとの逆適用範囲に限定され、それ以外の複合項に変数の置換や、項の代入などの影響はまったくない。従って、両方の規則を同時に逆適用することが可能である。

## 6.1 並列解析の定義

UPPUG において逆適用可能な複数の規則が、並列的に逆適用可能になるのは、各規則が UPPUG の制約を満たし、ある規則と他の規則とが逆適用位置において重なっても、それらの部分が規則それぞれの文脈インデックスが示す範囲内にあるからである。従って、各規則の文脈以外の部分を同時に書換えることにより、並列還元が実現できる。

**定義 6.1[並列還元]** $G = (F, T, V, P, s^{(n_s)}, \$)$ を UPPUG、$\xi \in (CTerm(F, V) \cup \{\$\})^+$ を任意の複合項列とする。二つの規則 $R_1 = [\alpha_1 \to \beta_1,\ (l_1, r_1)]$ と $R_2 = [\alpha_2 \to \beta_2,\ (l_2, r_2)]$ がそれぞれ位置 $i_1$, $i_2$ において、$\xi$ に対して並列逆適用可能であるとは、次の (1) または (2) が成り立つことを言う。

(1) $i_2 - i_1 \geq |\beta_1|$ かつ、ある $\gamma, \theta, \lambda \in (CTerm(F, V) \cup \{\$\})^+$ と代入 $\sigma_1$, $\sigma_2$ が存在し、
$$\xi = \gamma, \beta_1\sigma_1, \theta, \beta_2\sigma_2, \lambda$$
と書ける。

(2) $i_2 - i_1 \leq |\beta_1|$ かつ、ある $\gamma, \lambda, \alpha'_1, \beta'_1, \alpha'_2, \beta'_2, \delta', \delta_1, \delta_2 \in (CTerm(F, V) \cup \{\$\})^*$ と代入 $\sigma_1$, $\sigma_2$ が存在し、
$$\xi = \gamma, \beta'_1\sigma_1, \delta', \beta'_2\sigma_2, \lambda$$
$$R_1 = [\alpha'_1\delta_1 \to \beta'_1\delta_1,\ (l_1, r_1)]$$
$$R_2 = [\delta_2\alpha'_2 \to \delta_2\beta'_2,\ (l_2, r_2)]$$
$$\delta' = \delta_1\sigma_1 = \delta_2\sigma_2$$
$$|\delta'| \leq min\{r_1, l_2\}$$
を満たす。

ここで、$\eta$ を

(1) の場合: $\eta = \gamma, \alpha_1\sigma_1, \theta, \alpha_2\sigma_2, \lambda$

(2) の場合: $\eta = \gamma, \alpha'_1\sigma_1, \delta', \alpha'_2\sigma_2, \lambda$

とする時、$G$ において $\eta$ が $\xi$ から並列的な直接還元によって得られたと言い、

$$\xi \underset{G}{\Longleftarrow} \eta\ \{[R_1, i_1, \sigma_1],\ [R_2, i_2, \sigma_2]\}$$

または

$$\xi \underset{G}{\overset{\{[R_1, i_1, \sigma_1],\ [R_2, i_2, \sigma_2]\}}{\Longleftarrow}} \eta$$

と書く。

上記の定義を $m$ 個の規則に対する並列還元に拡張するのは容易である (但し、(2) の場合のように、規則の逆適用範囲が重なる時の記述が繁雑になる)。このような並列的直接還元を

$$\xi \underset{G}{\Longleftarrow} \eta\ \{[R_1, i_1, \sigma_1], \cdots, [R_m, i_m, \sigma_m]\}$$

または

$$\xi \underset{G}{\overset{\{[R_1, i_1, \sigma_1], \cdots, [R_m, i_m, \sigma_m]\}}{\Longleftarrow}} \eta$$

と書く。 □

UPPUG では、複合項列 $\xi$ に対して、複数の $m$ 個の規則が並列逆適用可能で、それらによる並列還元と、各々の規則が $\xi$ に逆適用可能で、それゆえ、各規則を順番に

逆適用することが可能となり、その直列還元とが、同値になることは次の補題により導くことができる。

**補題 6.1** $G = (F, T, V, P, s^{(n_s)}, \$)$ を UPPUG、$\xi, \eta \in (CTerm(F,V) \cup \{\$\})^+$ を複合項列、$R_1, \cdots, R_m$ を $P$ 中の規則、$i_1, \cdots, i_m$ を非負の整数、$\sigma_1, \cdots, \sigma_m$ を代入とする。この時、次の (1) と (2) が同値である。

(1) $\xi \overset{\{[R_1,i_1,\sigma_1],\cdots,[R_m,i_m,\sigma_m]\}}{\Longleftarrow} \eta$

(2) ある $\gamma_1, \cdots, \gamma_{m-1} \in (CTerm(F,V) \cup \{\$\})^+$ と、非負整数 $i'_1, \cdots, i'_m$ に対して、

$$\xi \overset{[R_1,i'_1,\sigma_1]}{\Leftarrow} \gamma_1 \Leftarrow \cdots \Leftarrow \gamma_{m-1} \overset{[R_m,i'_m,\sigma_m]}{\Leftarrow} \eta$$

**(証明概略)** $\{[R_1, i_1, \sigma_1], \cdots, [R_m, i_m, \sigma_m]\}$ が $\xi$ に並列逆適用可能であることと、各々の $[R_k, i_k, \sigma_k]$ $(1 \le k \le m)$ が $\xi$ に逆適用可能であることが同値であることは、並列逆適用可能性の定義から明らかである。また、(1) の並列還元と (2) の直列還元のいずれによっても $\eta$ が得られることは、補題 5.1 と同様の方法によって証明できる。 □

規則の左右の長さや、逆適用位置などは、直接還元の実行に当たって、容易に把握できるものである。そして、各規則には予め文脈インデックスを付随させ、簡単に共有することができる。従って、逆適用可能な規則の数に応じて配置される、しかもそれぞれの規則を逆適用する並列プロセッサ同士が、以上の情報を用いて、規則の文脈範囲以外の部分を同時に (並列に) 書換えることができる。

### 6.2 一意並列解析性

**定理 6.1[一意並列解析性]** $G = (F, T, V, P, s^{(n_s)}, \$)$ を UPPUG、$\eta \in (CTerm(F,V) \cup \{\$\})^+$ を複合項列とする。ある項 $t_1, \cdots, t_{n_s} \in Term(F,V)$ に対して、

$$\eta \overset{n}{\Leftarrow} \$s(t_1, \cdots, t_{n_s})\$$$

が成り立つと仮定する。この時にもし、

$$\{[R_1, i_1, \sigma_1], \cdots, [R_m, i_m, \sigma_m]\}$$

が $\eta$ に対して並列逆適用可能ならば、ある複合項列 $\xi \in (CTerm(F,V) \cup \{\$\})^+$ が存在して、

$$\eta \overset{\{[R_1,i_1,\sigma_1],\cdots,[R_m,i_m,\sigma_m]\}}{\Longleftarrow} \xi \overset{n-m}{\Leftarrow} \$s(t_1, \cdots, t_{n_s})\$$$

が成り立つ。

**(証明)** 補題 6.1 により、

$$\eta \overset{[R_1,i'_1,\sigma_1]}{\Leftarrow} \cdots \overset{[R_m,i'_m,\sigma_m]}{\Leftarrow} \xi$$

および

$$\eta \overset{\{[R_1,i_1,\sigma_1],\cdots,[R_m,i_m,\sigma_m]\}}{\Longleftarrow} \xi$$

が成り立つ。さらに、$\eta$ に対し、UPPUG の一意解析性 (定理 5.1) を $m$ 回適用することによって、

$$\eta \overset{[R_1,i'_1,\sigma_1]}{\Leftarrow} \cdots \overset{[R_m,i'_m,\sigma_m]}{\Leftarrow} \xi \overset{n-m}{\Leftarrow} \$s(t_1, \cdots, t_{n_s})\$$$

が得られる。従って、

$$\eta \overset{\{[R_1,i_1,\sigma_1],\cdots,[R_m,i_m,\sigma_m]\}}{\Longleftarrow} \xi \overset{n-m}{\Leftarrow} \$s(t_1, \cdots, t_{n_s})\$$$

となり、定理が成り立つ。 □

## 7 まとめ

本研究では、ユニフィケーション文法（UG）の書換規則に文脈インデックスを導入し、そして、UPG における書換規則の制約と同様の制約を加えることにより、一意並列解析可能ユニフィケーション文法（UPPUG）の定式化を行った。UPPUG は UPG のバックトラックなしに、解析が可能という一意解析可能性に加え、新たに並列還元が定義され、一意並列解析可能性も併せ持つ文法体系である。なおかつ、UPG と同じく万能な生成文法である。また、非終端記号に引数を持たせ、ユニフィケーション計算によって、情報のやり取りを簡潔に行うことができるので、自然言語、計算機言語などに対する文法の記述力が高められ、文法の設計も容易になる。

## 参考文献


[1] K. Morita, N. Nishihara, Y. Yamamoto and Z. Zhang: A hierarchy of uniquely parsable grammar classes and deterministic acceptors, *Acta Informatica*, **34**, 389-410, 1997.

[2] Y. Yamamoto and K. Morita: Two-dimensional uniquely parsable isometric array grammars, *International Journal of Pattern Recognition and Artificial Intelligence*, **6**, 301-313, 1992.

[3] J.E. Hopcroft and J.D. Ullman: Introduction to Automata Theory, Languages, and Computation, Addison-Wesley, Reading, Massachusetts, 1979.

[4] A.V. Aho and J.D. Ullman: The Theory of Parsing, Translation, and Compiling: Volume 1: Parsing, Prentice-Hall, 1972.

[5] G. Gazdar and C. Mellish: Natural language processing in LISP : an introduction to computational linguistics, Addison-Wesley Publishing Company, Menlo Park, CA, 1989.