

5.2 문맥자유문법의 문법나무와 유도순서

(정의 5.9) 문법나무(parse tree, syntactic tree)

문맥자유문법 $G = (N, T, P, S)$ 에서 $X \in N \cup T$ 의 **문법나무** $T_G(X)$ 를 X 가 뿌리인 **뿌리 나무**(rooted tree) $T_G(X) = (V, E, X)$ ¹⁾로 아래와 같이 정의한다.

(basis) $a \in T$ 에 대하여 뿌리가 a 인 **문법나무** $T(a) = (\{a\}, \emptyset, a)$ 다.

(recursion) $A \rightarrow X_1 X_2 \cdots X_n \in P$ (단 $A \in N, 1 \leq \forall i \leq n: X_i \in N \cup T$)에서,

$1 \leq \forall i \leq n: T_G(X_i) = (V_i, E_i, X_i)$ 가 X_i 를 뿌리로 하는 **문법나무**이면,

A 를 뿌리로 하는 새로운 **문법나무** $T_G(A) = (V_A, E_A, A)$ 는

$$(1) V_A = \{A\} \cup \bigcup_{i \in N_n} V_i,$$

$$(2) E_A = \{(A, X_i) | A \rightarrow X_1 X_2 \cdots X_n \in P\} \cup \bigcup_{i \in N_n} E_i \text{로 정의한다.}$$

(정의 5.10) 쓴 문법규칙열(rule string)을 표기하도록 유도(\Rightarrow)를 확장, 즉 \Rightarrow^π 단 $\pi \in P^*$.

(basis) $\alpha \Rightarrow^\epsilon \alpha$, For $\alpha \in V^*$.

(recursion) If $\alpha \Rightarrow^\pi \beta \wedge \beta \Rightarrow^r \gamma$, then $\alpha \Rightarrow^{\pi \cdot r} \gamma$. 단 $\pi \in P^*, r \in P$.

(정의 5.11) 문법 $G = (N, \Sigma, P, S)$ 에서 $S \Rightarrow^\pi x$ (단 $\pi \in P^*$)이면 **문장** $x \in T^*$ 에 대한 문법 나무 $T_G(S)$ 는 S 를 뿌리로 하는 문법나무 $T_G(S) = (V, E, S)$ 중 잎(leaves)이 문자열 x 이고 이때 사용된 규칙열이 π 인 경우이다.

$S \Rightarrow^\pi x \in T^*$ 이고 $\pi \in P^*$ 일 때, 사용된 문법규칙으로 subtree를 만들면 전체 문장 $x \in T^*$ 에 대한 트리를 만들 수 있고, 이를 파스트리라 부른다.

문맥자유문법의 **문장형태**는 일반적으로 여러 개에 년 터미널을 가지는데, 어떤 년 터미널이 유도에 **먼저** 참여할까 하는 것은 문맥자유-문법나무에 최종모양과는 관계가 **없다**²⁾. 따라서 문장형태에서 가장 왼쪽에 있는 년 터미널부터 다시 쓰기를 하는 왼쪽부터 고쳐 쓰기(leftmost derivation)과 오른쪽부터 고쳐 쓰기(rightmost derivation)에 두 가지 극단적인 경우를 생각할 수 있다. 이를 각각 \Rightarrow_{lm} 와 \Rightarrow_{rm} 로 표시한다.

(정의 5.12) 왼쪽부터 고쳐 쓰기(leftmost derivation): $\Rightarrow_{lm}^{\pi_L}, \pi_L \in P^*$.

(basis) $\alpha \Rightarrow_{lm}^\epsilon \alpha$, For $\alpha \in V^*$.

(recursion) If $\alpha \Rightarrow_{lm}^\pi x B \gamma$ and $x B \gamma \Rightarrow_{lm}^{B \rightarrow \beta} x \beta \gamma$, then $\alpha \Rightarrow_{lm}^{\pi \cdot B \rightarrow \beta} x \beta \gamma$.

단 $x \in T^*, B \in N, B \rightarrow \beta \in P, \pi \in P^*$.

(정의 5.13) 오른쪽부터 고쳐 쓰기(rightmost derivation): $\Rightarrow_{rm}^\pi, \pi_R \in P^*$.

(basis) $\alpha \Rightarrow_{rm}^\epsilon \alpha$, For $\alpha \in V^*$.

1) 뿌리나무 $T = (V, E, X)$ 에서 (1) V 는 vertex의 집합, (2) $E \subseteq V \times V$ 는 edge의 집합, (3) $X \in V$ 는 뿌리라 불리는 특별한 vertex이다. 그래프 $G = (V, E)$ 와 뿌리나무 $T = (V, E, X)$ 에 관한 기초적인 정의는 CS204 이산수학 참조.

2) Context-free 문법의 특성이다.

(recursion) If $\alpha \Rightarrow_{rm}^{\pi} \alpha Bz$ and $\alpha Bz \Rightarrow_{rm}^{B \rightarrow \beta} \alpha \beta z$, then $\alpha \Rightarrow_{rm}^{\pi \cdot B \rightarrow \beta} \alpha \beta z$.
 단 $z \in T^*$, $B \in N$, $B \rightarrow \beta \in P$, $\pi \in P^*$.

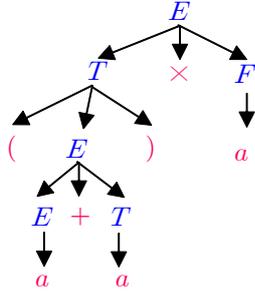
문장

(예 5.1) $G = (\{E, T, F\}, \{a, +, \times, (,)\}, P, E)$ 에서
 $P: E \rightarrow E + T \mid T \times F \mid a \mid (E)$
 $T \rightarrow T \times F \mid a \mid (E)$
 $F \rightarrow a \mid (E)$

$(a + a) \times a$ 에 관한 두 가지 유도과정

$E \Rightarrow_{lm}^{E \rightarrow T \times F} T \times F \Rightarrow_{lm}^{T \rightarrow (E)} (E) \times F \Rightarrow_{lm}^{E \rightarrow E + T} (E + T) \times F \Rightarrow_{lm}^{E \rightarrow a} (a + T) \times F$
 $\Rightarrow_{lm}^{T \rightarrow a} (a + a) \times F \Rightarrow_{lm}^{F \rightarrow a} (a + a) \times a.$

$\pi_L = (E \rightarrow T \times F) \cdot (T \rightarrow (E)) \cdot (E \rightarrow E + T) \cdot (E \rightarrow a) \cdot (T \rightarrow a) \cdot (F \rightarrow a) \in P^6.$



$E \Rightarrow_{rm}^{E \rightarrow T \times F} T \times F \Rightarrow_{rm}^{F \rightarrow a} T \times a \Rightarrow_{rm}^{T \rightarrow (E)} (E) \times a \Rightarrow_{rm}^{E \rightarrow E + T} (E + T) \times a$
 $\Rightarrow_{rm}^{T \rightarrow a} (E + a) \times a \Rightarrow_{rm}^{E \rightarrow a} (a + a) \times a.$

$\pi_R = E \rightarrow T \times F \cdot F \rightarrow a \cdot T \rightarrow (E) \cdot E \rightarrow E + T \cdot T \rightarrow a \cdot E \rightarrow a \in P^6.$

$\pi_R^{R3} = E \rightarrow a \cdot T \rightarrow a \cdot E \rightarrow E + T \cdot T \rightarrow (E) \cdot F \rightarrow a \cdot E \rightarrow T \times F \in P^6.$

$S \Rightarrow_{lm}^{\pi_L^1} xA\gamma \Rightarrow_{lm}^{B \rightarrow \beta} x\beta\gamma \Rightarrow_{lm}^{\pi_L^2} xy\gamma \Rightarrow_{lm}^{\pi_L^3} xyz, \quad \pi_L^1, \pi_L^2, \pi_L^3 \in P^*,$

단 $x \in T^*$, $\gamma \in (N \cup T)^*$, $B \rightarrow \beta \in P$, $y, z \in T^*$, $\beta \Rightarrow^* y$, $\gamma \Rightarrow^* z$.

$S \Rightarrow_{rm}^{\pi_R^1} \alpha Az \Rightarrow_{rm}^{B \rightarrow \beta} \alpha \beta z \Rightarrow_{rm}^{\pi_R^2} \alpha yz \Rightarrow_{rm}^{\pi_R^3} xyz, \quad \pi_R^1, \pi_R^2, \pi_R^3 \in P^*,$

단 $z \in T^*$, $\alpha \in (N \cup T)^*$, $B \rightarrow \beta \in P$, $y, x \in T^*$, $\beta \Rightarrow^* y$, $\alpha \Rightarrow^* x$.

이 때 $\pi_L = \pi_L^1(A \rightarrow \beta)\pi_L^2\pi_L^3 \in P^*$ 라 하고 $\pi_R = \pi_R^1(A \rightarrow \beta)\pi_R^2\pi_R^3 \in P^*$ 라 하면 π_L 과 π_R 를 각각 문장 x 에 대한 left parse(좌 분석)와 right parse(우 분석)라고 부른다. 문장 x 에 대한 좌 분석이나 우 분석이 정해지면 해당하는 문법트리도 하나로(unique) 정해진다.

5.2 문맥자유문법과 문법나무 그리고 유도순서

문맥자유문법의 문장형태는 일반적으로 여러 개에 낀 터미널을 가지는데, 어떤 낀 터미널이

3) π_R^R 의 위에 R 은 오른 파스 π_R 을 뒤집은 것(Reversal)을 뜻한다.

유도에 먼저 참여할까 하는 것은 파스트리에 최종모양과는 관계가 없다. 따라서 문장형태에서 가장 왼쪽에 있는 n 터미널부터 다시 쓰기를 하는 왼쪽부터 고쳐 쓰기(leftmost derivation)과 오른쪽부터 고쳐 쓰기(rightmost derivation)에 두 가지 극단적인 경우를 생각할 수 있다. 이를 각각 \Rightarrow_{lm} 와 \Rightarrow_{rm} ⁴⁾로 표시한다.

$$S \Rightarrow_{lm}^{\pi_1} xA\gamma \Rightarrow_{lm}^{B \rightarrow \beta} x\beta\gamma \Rightarrow_{lm}^{\pi_2} xy\gamma \Rightarrow_{lm}^{\pi_3} xyz, \quad \pi_1, \pi_2, \pi_3 \in P^*,$$

단 $x \in T^*$, $\gamma \in (N \cup T)^*$, $B \rightarrow \beta \in P$, $y, z \in T^*$, $\beta \Rightarrow_{lm}^{\pi_2} y$, $\gamma \Rightarrow_{lm}^{\pi_3} z$.

$$S \Rightarrow_{rm}^{\pi_1^R} \alpha Az \Rightarrow_{rm}^{B \rightarrow \beta} \alpha \beta z \Rightarrow_{rm}^{\pi_2^R} \alpha y z \Rightarrow_{rm}^{\pi_3^R} xyz, \quad \pi_1^R, \pi_2^R, \pi_3^R \in P^*,$$

단 $z \in T^*$, $\alpha \in (N \cup T)^*$, $B \rightarrow \beta \in P$, $y, x \in T^*$, $\beta \Rightarrow_{lm}^{\pi_2} y$, $\alpha \Rightarrow^* x$.

이 때 $\pi_L = \pi_L^1(A \rightarrow \beta)\pi_L^2\pi_L^3 \in P^\dagger$ 라 하고 $\pi_R = (\pi_R^1(A \rightarrow \beta)\pi_R^2\pi_R^3)^R \in P^\dagger$ 라 하면 π_L 과 π_R 를 각각 문장 x 에 대한 left parse(좌 분석)와 right parse(우 분석)라고 부른다. 문장 x 에 대한 좌 분석이나 우 분석이 정해지면 해당하는 문법트리도 정해진다.

문맥자유문법의 파싱(parsing)은 임의의 터미널 문자열 $x \in T^*$ 가 주어진 문법 G 에 문장이면 해당하는 파스트리를 만들어주고, 문장이 아니면 이를 알려주는 과정이다.

4) $\Rightarrow_{lm}, \Rightarrow_{rm} \subseteq P^*$ 이다.