LL 언어의 특징화
(Characterization of LL languages)

이 경 옥 †
(Gyung-Ok Lee)

요 약 주어진 문법이 LL 언어를 생성하는가의 문제를 LL 변환 가능성의 관점에서 조사한다. LL 변환 가능 문법을 특징짓기 위해 기존의 방법은 비결정적인(nondeterministic) 복잡한 파서의 생성을 필요로 한다. 이에 기존 방법으로는 LL 언어의 본질에 대한 직관적인 이해가 어렵다. 이 단점을 극복하고자 본 논문에선 파서의 생성없이 문법 자체의 분석을 통한 문법 유도(derivation)에 근거한 특징화를 제시한다. 새로운 특징화는 LL 언어의 본질에 대한 직관적인 이해를 도울 수 있다.
키워드 : LL 언어, LR 문법, LL 문법

Abstract The problem whether a given grammar G generates an LL language or not is investigated in respect of LL transformable grammars. The previous work involves a nondeterministic intricated parser construction for the characterization of LL transformable grammars. The method hence does not give the intuitive understanding of the essence of LL languages. This paper suggests a characterization of LL transformable grammars based on grammatical derivations instead of the complicated parser construction. The new characterization contributes to intuitive understanding of the essence of LL languages.
Key words : LL language, LR grammar, LL grammar

1. 서 론

주어진 문법이 생성하는 언어가 LL 언어인지를 판단하는 문제는 결정할 수 없는(undecidable) 문제로 알려졌다 [1]. 이에 LL 언어의 결정가능한(decidable) 부분 클래스에 관한 연구가 진행되었으며, 이에 대한 연구로는 주어진 문법이 제안된 변환을 거쳐서 LL 문법을 생성하는지의 여부에 따라서 대상 문법이 LL 언어를 생성하는지를 결정짓는 접근 방법이 있었다. 구체적으로는 휴리스틱에 의한 LL로의 변환 방법 [2,3,4]과 정형화된 방식으로 LL로의 변환 가능한 문법을 정의한 PLR [5], k-transformable [6,7], 확장된 k-transformable [8] 문법들에 관한 연구가 수행되었다. PLR 문법은 문법의 유도과정에 기반한 잘 특징지어진 문법인데 반해서, k-transformable 문법과 확장된 k-transformable 문법은 제안된 LL로의 변환 알고리즘을 적용해 본 후에 해당 문법인지를 파악할 수 있다. 이런 특징화 방법은 LL 변환 단계를 완전히 이해해야만, 각 문법 클래스의 본질을 파악할 수 있다는 단점이 있다. 더구나 이들이 제시한 변환 방법은 비결정적이기에 변환 가능한 문법 클래스에 대한 직관적 이해가 어렵다. 이러한 문제점을 해결하고자 본 논문에서는 k-transformable 문법에 대한 문법 유도에 근거한 특징화를 제시한다.

기존의 방법은 싸이클이 없는 다중 스택 파서를 얻기 위한 시행 착오에 의한 작업[6,7]이 필요한데 반해, 새로운 특징화는 휴리스틱 작업이 불필요하다. 따라서 제안되는 특징화는 직관적으로 LL 변환 가능 문법 클래스를 이해할 수 있게 돕는다. 본 논문에선 정형화를 위한 근본 아이디어를 k-transformable 문법에 우선적으로 적용하여 논문을 전개해 나가지만 확장된 k-transformable 문법에 대해서도 유사하게 적용할 수 있다.

2장에선 본 논문의 기본 배경을 언급하고, 3장에선 문법 유도에 근거한 특징화를 위한 정형식들을 소개한다. 4장에선 k-transformable 문법에 대한 문법 유도 특징화를 제시하고, 5장에선 결론을 맺는다.

· 본 논문은 부분적으로 2001년도 한신대학교 학술연구비 지원에 의하여 연구되었음.
† 정 회 원 : 한신대학교 정보통신학과 교수
golee@hanshin.ac.kr
논문접수 : 2000년 12월 26일
심사완료 : 2001년 10월 24일

2. 배 경

이 절에서는 본 논문의 전개를 위해 필요한 배경 지식들을 언급한다.

2.1 기본 정의들과 표기법

표기법과 기본 정의들은 [9,10]을 따르며 독자의 편의를 위해서 몇가지 정의들을 다시 언급한다.

심볼 G는 임의의 고정된 $G=(N,\Sigma,P,S)(V=N\cup\Sigma)$를 나타낸다. 심볼 k는 임의의 고정된 양수를 나타낸다. N에 속하지 않는 S'에 대해서 $G'=(N\cup\{S'\},\Sigma, P\cup\{S'\to S\$^k\},S')$는 G의 **확장된(augmented) 문법**이다. G는 S'에 의해 확장된 문법이고 **리듀스된(reduced) 문법**임을 가정한다. 또한 어떤 $k(k\geq 1)$에 대해 G는 LR(k)임을 가정한다. α를 스트링이라 하자. $k\alpha$는 $|\alpha|\leq k$이면 α와 동일하고 그 외의 경우엔 α의 처음 k개의 심볼들과 동일하다. 이때 $FIRST_k^G(\alpha)=\{k x|\alpha\Rightarrow^* x, x\in\Sigma^*\}$, $FOLLOW_k^G(\alpha)=\{k x|\ S'\Rightarrow^*\beta\alpha x, x\in\Sigma^*\}$로 정의된다. G가 문맥상 명확할 경우엔 G는 생략한다.

생존가능 스트링(viable prefix) α에 대한 **k-우문맥(right context)**은 $RC_k(\alpha)=\{k yz\ |\ S'\Rightarrow^*_{rm}\beta Bz\Rightarrow_{rm}\beta\gamma\delta z\Rightarrow^*_{rm}\beta\gamma yz, \beta\gamma=\alpha\}$로 정의한다.

$[A\to\alpha.\beta,u]$가 LR(k) 아이템이면 $L_k([A\to\alpha.\beta,u])=FIRST_k(\beta u)$이다. q가 LR(k) 아이템들의 집합이면 $L_k(q)=\{x\in FIRST_k(\beta u)|\ [A\to\alpha.\beta,u]\in q\}$이다.

$X\in V$이면 $FOLLOW_k(q,X)=\{x|\ [A\to\alpha.X\beta,u]\in q, x\in FIRST_k(\beta u)\}$이다. LR 아이템상에 클로져 전이 $\to^\varepsilon$은 $[A\to\alpha.B\beta,u]\to^\varepsilon[B\to.\gamma,v]$, $v\in FIRST_k(\beta u)$으로 정의된다. 전이 $\to^X$은 $[A\to\alpha.X\beta,u]\to^X[A\to\alpha X.\beta,u]$으로 정의된다. 전이 $\to$은 LR 아이템들상의 정의로부터 LR 아이템들의 집합상으로 확장된다.

$[A\to\alpha.X\beta,u]$에 대해 $X\in\Sigma$ 또는 $X\beta=\varepsilon$이면 **terminal 아이템**이다. q가 LR(k) 아이템들의 집합이면 $T(q)=\{I\in q\ |\ I$는 terminal 아이템이다$\}$이다.

2.2 Hammer의 방법

커버링 문법을 얻기위해서 Hammer는 LR 기계의 정보를 이용했다. LR 상태는 (A,α)의 형태로 이름이 붙여지고, Hammer에 의해 제시된 타당한 상태 분리 조건을 만족할 경우에 상태는 분리될 수 있다. 이때 변환 문법을 생성하는 생성자는 타당한 상태 분리중에서 선택을 하여 *SplittingSet*을 결정하여서 *T(G,SplittingSet)*의 문법을 얻는다.

타당한 상태 분리를 위한 정형식을 제시하기 전에 LR(k) 아이템들 상의 몇가지 필요한 정의를 보인다.

$U\in 2^{\Sigma^k}$라고 하자. $c=I_0\to^\varepsilon I_1\to^\varepsilon\cdots\to^\varepsilon I_{n-1}\to^\varepsilon I_n$($I_0$은 q의 essential 아이템이고 I_n은 terminal 아이템이다.)에 대해서 $L_k(I_n)\cap U\neq\emptyset$이면 c는 **q를 통과하는 U-체인**이다. $[A\to.\theta,u]$ 형태의 아이템은 **A-아이템**이고 $L_k(q,A)$는 $\{x\in FIRST_k(A\beta u)|[B\to\alpha.A\beta,u]\in q\}$이다.

다음은 체인과 문법 유도 과정의 관계를 보인다.

성 질 2 . 1 체 인 $[C\to\delta.A_0\zeta,r]\to^\varepsilon[A_0\to.A_1\zeta_1,r_0]\to^\varepsilon\cdots\to^\varepsilon[A_{n-1}\to.A_n\zeta_n,r_{n-1}]$가 존재한다면 유도과정 $Cr\Rightarrow_{rm}\delta A_0\zeta r\Rightarrow^*_{rm}\delta A_0z_0r\Rightarrow_{rm}\Rightarrow^*_{rm}\delta A_1\zeta_1z_0r$ $\delta A_{n-1}z_{n-1}\cdots z_2z_1z_0r\delta A_n\zeta_nz_{n-1}\cdots z_2z_1z_0r\Rightarrow^*_{rm}\Rightarrow_{rm}$ $\delta A_nz_nz_{n-1}\cdots z_2z_1z_0r$ (각 $i=0,1,\cdots,n-1$에 대해서 $r_i=kz_i\cdots z_0r$)가 존재한다. □

타당한 상태 분리에 대한 조건을 다음에서 보인다.

정의 2.1 (타당한 상태 분리) 상태 q에 대한 $(Base,A,Prd)$로의 상태 분리가 **타당(valid)**하기 위한 필요 충분 조건은 다음과 같다. q를 통과하는 각 $L_k(q,A)$-체인 $c=I_0,\cdots,I_n$에 대해서 아래의 성질 (i), (ii)를 만족하는 m이 존재한다.

(i) I_{m-1}은 A-아이템이다.

(ii) $H1(c)=\{I_0,\ldots,I_m\}, H2(c)=\{I_{m+1},\ldots,I_n\}$일 때 $Base=\bigcup_c H1(c)\cup\{I\ |\ L_k(I)\not\subseteq L_k(q,A), I\in q\}$, $Prd==\bigcup_c H2(c)$라고 하자. 이때 (식 1) $FOLLOW_k(Base,A)\cap FOLLOW_k(Prd,A)=\emptyset$이다. □

A가 좌순환이 아닌 경우엔 $FOLLOW_k(Prd,A)=\emptyset$이기에 (식 1)은 항상 성립한다. 따라서 성질 (ii)는 A가 좌순환인 경우에만 영향을 미친다. 정의 2.1내에서 $L_k(q,A)$는 **예상 언어**이다. 예상 언어는 성질 2.2와 같이 계산될 수 있으며 증명은 [11]을 참조하기 바란다.

성질 2.2 q에 대한 *(Base,A,Prd)*로의 상태 분리가 타당하면 $L_k(q,A)=L_k(Prd)$가 성립한다. □

정의 2.1에서의 타당한 이분 상태 분리는 타당한 다중 상태 분리로 확장될 수 있다.

2.3 L 관계

LR 아이템들의 클로져 전이 관계를 nonterminal들 사이의 관계로 단순화하여 표현할 수 있다. 이를 위해서 L 관계를 정의한다. $A\in N, r\in\Sigma^k, X\in V\cup\{\varepsilon\}, w\in\Sigma^k\cup\Sigma^{k-1}$라 하자. $(A,r)\ L\ (X,w)$은 다음의 조건이 성립함을 의미한다. $A\to X\alpha\in P$, $r\in FOLLOW_k(A)$, $X\in N\cup\{\varepsilon\}$인 경우엔 $w\in FIRST_k(\alpha r)$이고, $X\in\Sigma$인 경

우엔 $w \in FIRST_{k-1}(ar)$이다. r, w가 중요하지 않는 경우에 $(A, r)\ L\ (X, w)$를 대신해서 $A\ L\ X$로 표현하기도 한다. L-관계와 대응되는 그래프는 L-그래프이다. L-그래프내의 경로는 $(A_{i-1}, r_{i-1})L(A_i, r_i), i=1,\cdots,n$일 때 $(A_0, r_0), (A_1, r_1), \cdots, (A_{n-1}, r_{n-1}), (A_n, r_n)$으로 나타낸다. $h=(A_0, r_0), \cdots, (A_n, r_n)$을 L-그래프상의 경로라 하자. $A_n \in \Sigma$ 또는 $A_n = \varepsilon$이면 h는 **terminal 경로**이다. h가 terminal 경로일 때 $U \in 2^{\Sigma^k}$에 대해서 $FIRST_k(A_n r_n) \cap U \neq \emptyset$이면 h는 ***U*-경로**이다. LR 아이템들의 집합 q에 대해서 $[C \to \delta.A_0\zeta, r] \in q$, $r_0 \in FIRST_k(\zeta r)$이면 h는 ***q*를 통과하는 *U*-경로**이다.

U-체인과 U-경로의 관계는 성질 2.3에서 보여진다.

성질 2.3 q를 통과하는 U-체인 $[C \to \delta.A_0\zeta, r] \to^{\varepsilon} [A_0 \to .A_1\zeta_1, r_0] \to^{\varepsilon} \cdots \to^{\varepsilon} [A_{n-1} \to .A_n\zeta_n, r_{n-1}]$이 존재하고 $r_n \in FIRST_{k-1}(\zeta_n r_{n-1})$이면 q를 통과하는 U-경로 $(A_0, r_0), (A_1, r_1), \cdots, (A_n, r_n)$가 존재한다. 이 역도 성립한다. □

q를 통과하는 U-체인은 q를 통과하는 U-경로에 의해서 인수분해됨을 성질 2.3으로부터 알 수 있다. 따라서 체인에 바탕을 둔 정형식들은 L-그래프상의 경로로 도메인을 바꿈으로 간단하게 표현될 수 있다.

3. 좌순환 nonterminal들에 대한 정형식

문법 변환 방법을 서술한 시에 좌순환 nonterminal의 특수성은 정형식을 복잡하게 만든다. 본 절에선 $\to^{\varepsilon}$, $\Rightarrow_{rm}$, L 관계들을 A와 $R(R \subseteq FOLLOW_k(A))$로 속성을 부여하는 제약된 관계들을 제시한다. 이는 좌순환 nonterminal에 대한 기존의 복잡성을 제거하여 비좌순환 nonterminal인 경우와 유사하게 처리되도록 만든다.

3.1 $\to^{\varepsilon}_{A,R}$ 전이

정의 2.1에서 A가 좌순환 nonterminal이 아닌 경우엔 각 $L_k(q, A)$-체인에 대해서 A-아이템은 유일하고 (성질 1) $Prd = \bigcup c\ \{J \mid I_{m+1} \to^{*} J\}$의 완전한 클로져(complete closure) 성질을 만족한다. 이 성질은 정의 2.1을 간략하게 만든다. 그러나 A가 좌순환인 경우엔 I_{m+1}을 기준으로 체인이 분리될 때 (성질 1)이 항상 성립하는 것은 아니다. 이에 이 절에서는 $\to$관계를 A와 R로 제약을 가해서 완전한 클로져 성질이 성립하도록 만들고자 한다.

$A \in N, R \subseteq FOLLOW_k(A)$라 하자. 또한 $[C \to \alpha.B\beta, u] \to^{\varepsilon} [B \to .\gamma, v]$라 하자. 이때 $B=A$이고 $v \in R$이면 $[C \to \alpha.B\beta, u] \to^{\varepsilon}_{A,R} [B \to .\gamma, v]$로 정의한다. 그리고 $q^{A,R}$과 $q^{A,R}_{\varepsilon}$을 각각 $\{[A \to .\theta, r] \mid A \to \theta \in P, r \in R\}$과 $\{J \mid I \to^{*\varepsilon}_{A,R} J,\ I \in q^{A,R}\}$로 정의하자. 이때 R을 $\bigcup_c LA_k(I_{m+1})$로 정의하면 $q^{A,R}_{\varepsilon}$는 Prd와 같음을 알 수 있다.

3.2 제약을 가한 유도 과정 $\Rightarrow_{A,R}$

$\to^{\varepsilon}_{A,R}$과 같은 취지로 유도 과정 $\Rightarrow_{rm}$에 제약을 가하고자 한다. $p = B \to X\delta$이고 $Ar \Rightarrow^{*}_{rm} \gamma Bzr \Rightarrow^{p}_{rm} \gamma X\delta zr$, $r \in R, z \in \Sigma^{*}$라 하자. 이때 $\gamma Bzr \Rightarrow^{p}_{A,R} \gamma X\delta zr$은 $\gamma \neq \varepsilon$이거나 $X \neq B$이면 항상 성립하고, $\gamma = \varepsilon$이고 $X = B$이면 $FIRST_k(\delta zr) \cap R \neq \emptyset$인 경우에 성립한다.

$\beta \neq \varepsilon$이면 $q^{A,R}_{\beta} = \{J \mid I \to^{*\beta} J,\ I \in q^{A,R}\}$로 정의한다. 다음의 성질은 $q^{A,R}_{\beta}$와 $\Rightarrow_{A,R}$의 관계를 보여준다.

성질 3.1 [11] $[B \to \delta.\zeta, u] \in q^{A,R}_{\beta}$이 존재하면 $Ar \Rightarrow^{*}_{A,R} \gamma Bzr \Rightarrow_{A,R} \gamma\delta\zeta zr (\gamma\delta = \beta)$이 G상에 존재한다. 또한 이 역도 성립한다. □

예상 언어는 상태 개념을 도입함 없이 표현될 수 있다. $L_k(A, R) = \{u \mid Ar \Rightarrow^{*}_{A,R} xr,\ u = k\,xr, r \in R\}$로 정의하면 다음과 같은 성질을 얻을 수 있다.

성질 3.2 [11] q에 대한 $(Base, A, Prd)$로의 상태 분리가 타당하다고 하자. q를 통과하는 각 $L_k(q, A)$-체인, $c = I_0, \cdots, I_n$에 대해 정의 2.1에 따라 $H1(c) = \{I_0, \cdots, I_m\}$, $H2(c) = I_{m+1}, \cdots, I_n\}$로 분리된 경우에 $R = \bigcup_c LA_k(I_{m+1})$로 정의하자. 이때 $L_k(q, A) = L_k(A, R)$이다. □

상태 q에 대해서 (1) $L_k(q) = U$이고 (2) (i) $\alpha = \varepsilon$일 때 $q = \{I \mid I \in q^{A,R}_{\varepsilon}\}$이고 (ii) $\alpha = X\beta (X \in V)$일 때 $q = \{I \mid J \in q^{A,R}_{\varepsilon},\ J \to^{X} K \to^{*\beta} I,\ L_k(I) \cap U \neq \emptyset\}$이면 q는 $q^{A,R}_{\alpha,U}$로 표기한다.

3.3 제약을 가한 L 관계

$(B, v)\ L\ (X, w)$라 하자. 이때 $X \neq A$ 또는 $w \notin R$이면 $(B, v)L_{A,R}(X, w)$라고 정의한다. 그리고 각 $i = 0, 1, \cdots, n-1$에 대해 $(X_i, w_i)L_{A,R}(X_{i+1}, w_i)$이면 $h = (X_0, w_0), (X_1, w_1), \ldots, (X_n, w_n)$는 **$(A, R)$-경로**라고 정의한다.

4. k-transformable 문법의 문법적 특징화

이 절에선 k-transformable 문법의 문법 유도에 기반한 특징화를 제시하고자 한다. 이를 위해 L 관계를 사용하여 2.2절의 타당한 상태 분리 정형식과 밀접히 관련된 $\Pi^{k-trans}$ 관계를 정의한다.

4.1 $\Pi^{k-trans}$ 관계

정의 4.1 ($(A, R, \alpha, U)\Pi^{k-trans}(B, W)$) $A, B \in N, R, U, W \in 2^{(\Sigma \cup \{\$\})^k}$, $\alpha \in V^+$라 하자. $(A, R, \alpha, U)\Pi^{k-trans}(B, W)$이기 위한 필요 충분 조건은 다음과 같다. $q_{\alpha,U}^{A,R}$를 통과하는 각 $L_k(q_{\alpha,U}^{A,R}, B)$-경로, $h=(A_0, r_0), (A_1, r_1), \ldots, (A_n, r_n)$에 대해 $(A_m, r_m), \ldots, (A_n, r_n)$이 (B, W)-경로이고 (식 2) $W=\bigcup_h \{r_m\}$인 $m(0 \le m \le n)$이 존재한다. (식 2)에서 h는 $q_{\alpha,U}^{A,R}$를 통과하는 모든 $L_k(q_{\alpha,U}^{A,R}, B)$-경로들을 지칭한다. □

다음 보조정리는 $\Pi^{k-trans}$ 관계와 타당한 상태 분리의 관계를 보인다.

보조정리 4.1 $(A, R, \alpha, U)\Pi^{k-trans}(B, W)$이면 $q_{\alpha,U}^{A,R}$에 대한 $(Base, B, q_c^{B,W})$로의 상태 분리는 타당하다. 이 역도 성립한다. □

정의 4.1에서 U-경로가 싸이클을 포함하더라도 [6]에서의 타당한 상태 분리를 찾을 시에 관련 체인에서 아이템들을 유한 번만 방문하는 원리와 같은 원리로 경로 상의 노드들은 유한 번 방문된다.

4.2 특징화

본 절에서는 $\Pi^{k-trans}$ 관계를 이용한 LL(k) 문법으로의 커버링 변환을 제시한다. 우선 LL 성질을 위한 $\Pi^{k-trans}$의 부분 관계를 정의한다.

정의 4.2 ($\Pi_{LL}^{k-trans}$) $\Pi_{LL}^{k-trans}$은 $\Pi^{k-trans}$의 부분 관계로 서로 다른 관계 $(A, R, \alpha, U)\Pi_{LL}^{k-trans}(B, W)$, $(A, R, \alpha, U)\Pi_{LL}^{k-trans}(C, X)$에 대해 $L_k(B, W) \cap L_k(C, X) = \emptyset$이어야 한다. □

$G, \Pi_{LL}^{k-trans}$가 주어졌을 때 $T^{k-trans}(G, \Pi_{LL}^{k-trans})$는 다음의 알고리즘에 따라 생성된다.

알고리즘 1 ($T^{k-trans}(G, \Pi_{LL}^{k-trans}) = (N^{k-trans}, \Sigma, P^{k-trans}, S^{k-trans})$의 생성)

입력 : $G, \Pi_{LL}^{k-trans}$

출력 : 이 알고리즘이 성공적으로 끝나면 $T^{k-trans}(G, \Pi_{LL}^{k-trans})$가 생성된다.

방법

1. 다음과 같이 초기화를 한다.

$S^{k-trans} = [S, \{\$^k\}, \varepsilon, FIRST_k(S\$^k)]$; $N^{k-trans} = \{S^{k-trans}\}$;

$P^{k-trans} = \emptyset$.

2. repeat

각 $[A, R, \alpha, U] \in N^{k-trans}$에 대해서

$Z = \{x \in L_k(B, W) | (A, R, \alpha, U)\Pi_{LL}^{k-trans}(B, W)\}$라 하자. 이때 다음의 단계들로 프러덕션을 생성한다.

(a) Type1. $P^{k-trans} = P^{k-trans} \cup \{[A, R, \alpha, U] \to a[A, R, \alpha a, V]\}$

여기서 $V = \{k{:}zwr | Ar \Rightarrow^*_{A,R} \beta Bwr \Rightarrow_{A,R} \beta\gamma a\delta wr \Rightarrow^*_{A,R} \beta\gamma azwr,$

$\beta\gamma = \alpha$, $k{:}azwr \in U$,

$k{:}azwr \notin Z$, $r \in R\}$로 정의되며 $V \neq \emptyset$이다.

(b)Type2. $P^{k-trans} = P^{k-trans} \cup \{[A, R, \alpha, U] \to [A, R, \beta B, V]\}$

여기서 $\alpha = \beta\gamma$이고 $V = \{k{:}wr | Ar \Rightarrow^*_{A,R} \beta Bwr \Rightarrow_{A,R} \beta\gamma wr$, $k{:}wr \in U$, $k{:}wr \notin Z$, $r \in R\}$로 정의되며 $V \neq \emptyset$이다.

(c) Type 3. $P^{k-trans} = P^{k-trans} \cup \{[A, R, \alpha, U] \to [B, W, \epsilon, V][A, R, \alpha B, W]\}$

여기서 $(A, R, \alpha, U)\Pi_{LL}^{k-trans}(B, W)$, $V = L_k(B, W)$이다.

(d) Type 4. $P^{k-trans} = P^{k-trans} \cup \{[A, R, A, U] \to \varepsilon\}$

여기서 $R \cap U \neq \emptyset$이고 $\alpha = A$이다.

until ($P^{k-trans}$는 변하지 않는다) □

이 알고리즘은 G에 따라 무한 수행을 하기도 한다. 이를 방지하기 위해서 아래와 같은 특정 성질을 갖는 nonterminal을 정의한다.

(1) $[A, R, \alpha, U] \in N^{k-trans}$이고 $\alpha = \beta\gamma(\gamma \neq \varepsilon)$라 하자. 이때

(i) $r \in R, w = k{:}zr(w \in W)$, $k{:}yw \in U$인 조건하에

$Ar \Rightarrow^*_{A,R} \beta Bzr$, $Bw \Rightarrow^*_{B,W} \gamma yw \Rightarrow^*_{B,W} xyw$이 존재하고

(ii) $V \subseteq \{k{:}yzr | Ar \Rightarrow^*_{A,R} \beta\gamma zr \Rightarrow^*_{A,R} \beta yzr$, $r \in R$, $k{:}zr \in U\}$인 $(A, R, \beta, V)\Pi^{k-trans}(B, W)$가 존재한다고 하자. 이때 $[A, R, \alpha, U]$는 **나눌 수 있다(divisible)**라고 한다. 그 밖의 경우엔 **나눌 수 없다(indivisible)**라고 한다.

(2) $[A, R, \alpha, U] \in N^{k-trans}$이고 $\alpha = X_1 \cdots X_n$라 하자. 이 때 $q_i = \{[B \to \gamma.\delta, k{:}zr] \mid r \in R$, $k{:}yzr \in U, \beta\gamma = X_1 \cdots X_i$, $\zeta = X_{i+1} \cdots X_n$인 조건하에 $Ar \Rightarrow^*_{A,R} \beta Bzr \Rightarrow_{A,R} \beta\gamma\delta zr \Rightarrow^*_{A,R} \beta\gamma\zeta yzr$이 존재한다. $\}$로 정의하자. 각 $q_0, q_1, \cdots, q_n$에 대해서 $q_i = q_j$ $(0 \le i < j \le n)$이고 $q_0, q_1, \cdots, q_n$의 모든 원소가 다르다면 $q_i, \cdots, q_j$를 **루프(loop)**라 한다. 같은 루프에 대해 $q_i, \cdots, q_j$이 루프인 i에 대해 서로 다른 값이 존재한다면 $[A, R, \alpha, U]$은 cyclic nonterminal이다.

$[A, R, \alpha, U]$이 나눌 수 없는 cyclic nonterminal인 경우에 $[A, R, \alpha, U]$내의 루프를 포함하는 cyclic nonterminal이 무한하게 존재하게 된다. 따라서 알고리즘 1을 수행 시에 나눌 수 없는 cyclic nonterminal이 발견되면 수행과정을 멈추어야 한다.

$T^{k-trans}(G, \Pi_{LL}^{k-trans})$내의 나눌 수 없는 cyclic nonterminal의 존재 여부는 $\Pi_{LL}^{k-trans}$에 의해서 결정된다. $\Pi_{LL}^{k-trans}$에 대응되는 상태 분리 집합을 *SplittingSet*이라고 하면 Hammer에 의한 변환 문법 *T(G,Splitting Set)*의 유한성 여부는 $T^{k-trans}(G, \Pi_{LL}^{k-trans})$내의 나눌 수 없는 cyclic nonterminal의 존재 여부와 동치이다.

따라서 알고리즘 1에 따라 변환될 수 있는 문법들은 k-transformable 문법이다.

정리 4.1 G에 대한 알고리즘 1이 성공적으로 끝나는 $\Pi_{LL}^{k-trans}$가 존재하는 것은 G가 k-transformable인 것과 동치이다. □

$P^{k-trans}$로부터 $P\cup\{\varepsilon\}$로의 동형 함수 $h^{k-trans}$를 다음과 같이 정의한다. r이 $[A, R, \beta\gamma, U] \rightarrow [A, R, \beta B, V]$이면 $h^{k-trans}(r) = B \rightarrow \gamma$이고 그 외에는 $h^{k-trans}(r) = \varepsilon$이다. 보조정리 4.3은 G와 $T^{k-trans}(G, \Pi_{LL}^{k-trans})$의 관계를 보이며 이에 대한 증명은 [12]를 참조하기 바란다.

보조정리 4.3 $G, \Pi_{LL}^{k-trans}$에 대하여 알고리즘 1이 성공적으로 끝났다고 가정하자. 만일 $S \Rightarrow_{rm}^{\pi} x$가 G상에 존재한다면 $h^{k-trans}(\pi_T) = \pi^R$을 만족하는 $[S, \{\$^k\}, \varepsilon, FIRST_k(S^k)] \Rightarrow_{lm}^{\pi_T} x$가 $T^{k-trans}(G, \Pi_{LL}^{k-trans})$상에 존재한다. 그 역으로 $[S, \{\$^k\}, \varepsilon, FIRST_k(S\$^k)] \Rightarrow_{lm}^{\pi_T} x$가 $T^{k-trans}(G, \Pi_{LL}^{k-trans})$상에 존재한다면 $h^{k-trans}(\pi_T) = \pi^R$을 만족하는 $S \Rightarrow_{rm}^{\pi} x$가 G상에 존재한다. □

알고리즘 1에 의해서 변환할 수 있는 문법들을 보조정리 4.4 (조건 1)로 특징화한 수 있으며 이에 대한 증명은 공간 제약상의 이유로 생략하며 [12]를 참조하기 바란다.

보조정리 4.4 G에 대해 알고리즘 1이 성공적으로 끝나는 $\Pi_{LL}^{k-trans}$이 존재하는 것은 (조건 1)이 성립하는 것과 동치이다.

(조건 1) G에 의해 결정되는 상수 n이 존재하여 다음을 만족한다. $\alpha, |\alpha| \geq n$가 G의 존속가능 스트링이고 $v \in RC_k(\alpha)$인 경우엔 $\alpha = \beta\gamma$, $|\gamma| \leq n$, $v \in RC_k^{B,W}(\gamma)$인 조건하의 B, W, γ가 존재하여, $S' \Rightarrow_{rm}^{*} \beta\gamma z \Rightarrow_{rm}^{*} \beta yz$, $k\,yz \in L_k(B, W)$이 존재할 때마다 $w = kz''(w \in W), z'z'' = z$인 $S' \Rightarrow_{rm}^{*} \beta Bz''$, $Bw \Rightarrow_{B,W}^{*} \gamma z'w \Rightarrow_{B,W}^{*} yz'w$이 존재한다. □

보조정리 4.4로부터 k-transformable 문법은 다음과 같이 특징지어질 수 있다.

정리 4.1 (**k-transformable 문법의 특징화**) G가 k-transformable일 필요 충분 조건은 (조건 1)이 참인 것이다. □

이 특징화의 장점은 (조건 1)에서는 다중 스택 기계의 복잡한 개념의 도입없이 새로 도입된 것은 유도 관계 $\Rightarrow_{A,R}$ 뿐이라는 것이다.

5. 결 론

본 논문에서는 기존 연구에서의 상태 분리에 관한 정형식을 L 관계를 이용한 $\Pi^{k-trans}$ 관계로 표현했고 $\Rightarrow_{A,R}$ 관계를 도입했다. 이를 바탕으로 k-transformable 문법에 대한 문법 유도 형태의 특징화를 제시했다.

본 논문의 기대 효과로 기존의 복잡한 파서 생성을 위한 정형식에 대한 이해없이 LL로의 커버링이 존재하는 변환 가능 문법을 직관적으로 알 수 있으며, 이는 LL 언어에 대한 직관적인 특징화를 가능하게 한다. 이로부터 LL 변환에 관한 새로운 아이디어의 정형적 접근을 수월하게 하여, 실용적으로 효율적인 LL 문법 변환의 개발에 기여하리라고 기대된다.

참 고 문 헌

[1] Rosenkrantz, D.J. and Stearns, R. E., "Properties of Deterministic Top-Down Grammars," Information and Control, Vol. 17, pp. 226-256, 1970.

[2] Foster, J.M., "A Syntax Improving Program," Comput. J., Vol 11, pp. 31-34, 1968.

[3] Wood, D., "The theory of left factored languages:PartI," The Computer Journal, Vol. 12, pp. 349-356, 1969.

[4] Wood, D., "The theory of left factored languages: PartII," The Computer Journal, Vol. 13, pp. 55-62, 1970.

[5] Soisalon-Soininen, E. and Ukkonen, E., "A Method for Transforming Grammars into LL(k) Form," Acta Informatica Vol. 12, pp. 338-369, 1979.

[6] Hammer, M., A New Grammatical Transformation into deterministic top down form. p. 301, MIT, Mass., Project MAC Technical Report TR-119, 1974.

[7] Hammer, M. "A New Grammatical Transformation into LL(k) form," Proc. of Sixth Annual ACM Symposium on Theory of Computing, pp. 266-275, 1974.

[8] 이경옥, 최광무, "LL(k) 커버링 문법의 확장", 정보과학회 논문지(B) 26권 8호, pp. 1028-1038, 1999.

[9] Aho, A. V. and Ullman, J. D., The Theory of Parsing, Translation and Compiling, vols.1 2. p. 1002, Englewood Cliffs, NJ:Prentice-Hall 1972, 1973.

[10] Sippu, S. and Soisalon-Soininen, E., Parsing

Theory, vol I, II, p. 228, p.426, Springer-Verlag Berlin Heidelberg, 1990.
[11] Lee, Gyung-Ok, "A grammatical derivation approach over k-transformable grammars," Journal of Hanshin Information and Science Research Institute, 2000.
[12] Lee, Gyung-Ok, On LL-to-LR Covering Languages, p. 74, Ph. D. Thesis, KAIST, 1999.

이 경 옥
1990년 2월 서강대학교 전자계산학과 졸업(학사). 1992년 8월 한국과학기술원 전산학과 졸업(석사). 2000년 2월 한국과학기술원 전산학과 졸업(박사). 현재 한신대학교 정보통신학과 조교수. 관심분야는 프로그래밍 언어론과 컴파일러 구성론등