229. *Covering-Languages of Grammars*

By Takumi KASAI
Research Institute for Mathematical Sciences, Kyoto University

(Comm. by Kinjirô KUNUGI, M. J. A., Dec. 13, 1971)

1. Introduction.

Two derivation trees (phrase-markers) are called *congruent* in [1] if merely by relabelling of the nonterminal nodes they may be made the same. A *marker* is an equivalence class of congruent derivation trees. In this note we introduce a new type of language, called a *covering language*, which can describe the set of markers generated by a context-free grammar. The intrinsic structure of a context-free grammar G is characterized by the covering language $K(G)$ of G.

Let $G=(N, \Sigma, P, S)$ be a context-free grammar with the set of nonterminal symbols N, the set of terminal symbols Σ, the set of productions P and the initial symbol S. Each production π is usually expressed in a unique way in the following canonical form

$$\pi : X \rightarrow t_0 Y_1 t_1 \cdots t_{n-1} Y_n t_n$$

where X and Y_i $(1 \leq i \leq n)$ are nonterminal symbols and the t are possibly empty terminal words. The integer $n \geq 0$ determines the number of occurrences of nonterminal symbols at the right side of the production π and is said to be the *rank* of π. The rank of a production π is denoted by $\sigma_P(\pi)$. For each production $\pi : X \rightarrow t_0 Y_1 t_1 \cdots Y_n t_n$, let $\langle t_0, t_1, \cdots, t_n \rangle$ be an abstract symbol. We shall call this the *form* of π and the integer n is said to be the *rank* of this form. The form of π will be denoted by $f(\pi)$ and the set of all forms of the productions in P will be denoted by $f(P)$, i.e. $f(P)=\{f(\pi) \mid \pi \text{ in } P\}$. We extend f to a length preserving homomorphism $f : P^* \rightarrow \{f(P)\}^*$ by defining $f(\varepsilon)=\varepsilon$ and $f(\pi_1 \cdots \pi_k)=f(\pi_1) \cdots f(\pi_k)$.

The notation $x \overset{\alpha}{\Longrightarrow} y$ or $\alpha : x \Longrightarrow y$ means that there exists a leftmost derivation

$$D : x = x_0 \overset{\pi_1}{\Longrightarrow} x_1 \overset{\pi_2}{\Longrightarrow} \cdots \overset{\pi_n}{\Longrightarrow} x_n = y$$

such that $\alpha = \pi_1 \pi_2 \cdots \pi_n$, where in the transition from x_i to $x_{i+1} (0 \leq i < n)$ the production π_i is applied. The word $\pi_1 \pi_2 \cdots \pi_n$ is called the *associate* of D and $f(\pi_1 \pi_2 \cdots \pi_n)$ is called the *form* of D.

In this paper, unless stated otherwise, by "grammar" we shall mean context-free grammar and by "derivation" we shall mean leftmost derivation.

Given a grammar $G=(N, \Sigma, P, S)$, let