96. An Improved Proof for a Theorem of N. Chomsky

By Masako TAKAHASHI
Department of the Foundations of Mathematical Sciences,
Tokyo University of Education

(Comm. by Zyoiti SUETUNA, M. J. A., June 10, 1969)

A context-free grammar is said to be self-embedding if and only if it is reduced and contains a derivation of the form $\xi \overset{*}{\Rightarrow} u\xi v$, where ξ is a variable and u, v are some non-ε words. It is known that

Theorem. *A language L is regular if and only if there is a non-self-embedding grammar generating L.*

This theorem was first presented in Chomsky [1] with a lengthy proof. Later a simplified proof was given in Chomsky [2]. In this note, the proof is improved by introducing some equivalence classes of the variables.

Our notations generally follow Ginsburg [3] with the additional convention that the variables are denoted by Greek small letters and words by Latin small letters.

1. Preliminaries. If a language L is regular, there exists a one-sided-linear grammar G such that $L(G)=L$, which is of course not self-embedding. So it is enough if we prove the regularity of $L(G)$ for any given reduced non-self-embedding (n.s.e) grammar $G=(V,\Sigma,P,\sigma)$.

Write $\xi \supseteq \eta$ whenever there exists at least one derivation of the form $\xi \overset{*}{\Rightarrow} u\eta v$, where $u, v \in V^*$. The relation $\supseteq$ is reflexive and transitive.

Write $\xi \equiv \eta$ if and only if $\xi \supseteq \eta$ and $\eta \supseteq \xi$. The relation $\equiv$ is an equivalence relation between variables.

Let $V(\xi)$ be the equivalence class under $\equiv$ containing ξ: $V(\xi) =\{\eta \mid \eta \in V-\Sigma,\ \eta \equiv \xi\}$, introducing a partial ordering between the equivalence classes as follows: $V(\xi) \geqslant V(\zeta)$ if and only if $\xi \supseteq \zeta$. Let $U(\xi)$ be the union of all $V(\zeta)$ such that $V(\xi) \gneqq V(\zeta)$ and Σ: $U(\xi) =\{\zeta \mid \zeta \in V-\Sigma,\ \xi \supseteq \zeta,\ \xi \not\equiv \zeta\} \cup \Sigma$.

2. Lemma. *Continuing with a reduced n.s.e grammar suppose that $\xi_1 \equiv \xi_2 \equiv \xi_3 \equiv \xi_4$ and $\xi_1 \overset{*}{\Rightarrow} s\xi_2 t$, $\xi_3 \overset{*}{\Rightarrow} u\xi_4 v$. Then either $t=v=\varepsilon$ or $s=u=\varepsilon$.*

Proof. By the definition of $\equiv$, there exist the derivations of the form $\xi_2 \overset{*}{\Rightarrow} s'\xi_3 t'$ and $\xi_4 \overset{*}{\Rightarrow} u'\xi_1 v'$. Then

$$\xi_1 \overset{*}{\Rightarrow} ss'uu'\xi_1 v'vt't.$$

In this derivation, assume that either t or v is a non-ε word. Then $ss'uu'$ must be ε, since the grammar is n.s.e. So $s=u=\varepsilon$, completing the proof.

3. Proof of the theorem, concluded. Let $P(\xi)$ be the set of all production rules in P whose left-hand sides belong to $V(\xi)$. By the lemma, the grammar $G(\xi)=(V(\xi)\cup U(\xi), U(\xi), P(\xi), \xi)$ is one-sided-linear. Hence, we may write a regular expression on $U(\xi)$ representing $L(G(\xi))$. If in particular $V(\xi)$ is minimal with respect to the ordering $\geqslant$, then $U(\xi)=\Sigma$, so that the regular expression is on Σ. Substituting one by one the expressions written for the variables of lower classes into those written for the variables of higher classes, for each variable we get a regular expression on Σ. The expression thus obtained for σ coincides with $L(G)$. This proves that $L(G)$ is regular.

References

[1] N. Chomsky: On certain formal properties of grammars. Information and Control, **2**, 137–167 (1959).
[2] ——: A note on phrase structure grammars. Information and Control, **2**, 393–395 (1959).
[3] S. Ginsburg: The Mathematical Theory of Context-free Languages. McGraw-Hill Book Company (1966).