INFORMATION AND CONTROL 7, 366–368 (1964)

Inherent Ambiguity of Minimal Linear Grammars

MAURICE GROSS

C.N.R.S. Laboratoire de Calcul Numérique de l'Institut Blaise Pascal, Paris

We give an example of minimal linear language, all of whose minimal linear grammars are ambiguous; this language is not ambiguous in the class of linear context-free languages.

Parikh has shown that there are context-free languages which are inherently ambiguous; he gave an example of context-free language (in fact linear), all of whose context-free grammars are ambiguous. Chomsky conjectured that there are languages inherently ambiguous relative to a class of grammars A, but not ambiguous in a class B, with $A \subset B$. We give here such a type of language where A is the class of minimal linear grammars, and B the class of linear grammars.

According to Chomsky and Schützenberger, a minimal linear language is generated by a context-free grammar of the particular form:

$$\left| \begin{array}{l} \{S \to f_i\, Sg_i : \quad 1 \leqq i \leqq N\} \\ S \to c \end{array} \right|$$

Our example is the language

$$L = \{a^m ca^n : m \geqq n \geqq 0\}$$

It is a minimal linear language, since the following grammar generates it:

$$\left| \begin{array}{l} S \to aSa \\ S \to aS \\ S \to c \end{array} \right| \tag{1}$$

We show that any minimal linear grammar G which generates L is ambiguous.

It follows from the definition that any such G that can generate L is of the form:

$$\left| \begin{array}{l} \{S \to a^{p_i} Sa^{q_i} : \quad 1 \leqq i \leqq N\} \\ S \to c \end{array} \right| \tag{2}$$

We have $p_i \geqq q_i$. Otherwise (2) would generate sentences of the type $a^m ca^n$ with $n > m$ which are not in L. We have, then, the three following types of rules:

$$\text{(i)} \qquad S \to a^{p_k} S \qquad p_k > 0$$

$$\text{(ii)} \qquad S \to a^{p_j} S a^{q_j} \qquad p_j \geqq q_j > 0$$

$$\text{(iii)} \qquad S \to c$$

The grammars (2), so constrained, generate sentences of the form: $a^m ca^n$, $m \geqq n \geqq 0$; in particular, they have to generate the sentences ac and aca. This entails that one of the rules of type (i) must be $S \to aS$ and one of the rules of type (ii) must be $S \to aSa$.

Then any grammar G has in fact the form:

$$\left| \begin{array}{l} S \to aSa \\ S \to aS \\ S \to c \\ \{S \to a^{p_j} S a^{q_j} : \quad p_j \geqq q_j \geqq 1\} \end{array} \right| \qquad (3)$$

The effect of the rules $\{S \to a^{p_j} S a^{q_j} : p_j \geqq q_j \geqq 1\}$ can be obtained from q_j applications of the rule $S \to aSa$ and $p_j - q_j$ applications of the rule $S \to aS$. Suppressing these rules in (3) decreases the degree of ambiguity of sentences of L; so the grammar (1) is the minimal linear grammar which generates L with a minimum of ambiguity.

A particular sentence $a^m ca^n$ is generated by n applications of rule $S \to aSa$, $m - n$ applications of rule $S \to aS$, and at the end of the derivation, by application of rule $S \to c$; the ambiguity of a sentence results from the order in which rules $S \to aSa$ and $S \to aS$ are applied. The degree of ambiguity of the sentence $a^m ca^n$ is then $\binom{m}{n}$; it is not bounded when the length of the sentence increases.

The language L is not inherently ambiguous in the class of linear grammars: the following grammar generates it unambiguously:

$$\left| \begin{array}{l} S \to aSa \\ S \to aT \\ T \to aT \\ S \to c \\ T \to c \end{array} \right|$$

Acknowledgment

I am indebted to Professor M. P. Schützenberger for important suggestions.

Received: February 4, 1964

References

Chomsky, N. (1963), Formal properties of grammars. In "Handbook of Mathematical Psychology," D. Luce, E. Bush, and E. Galanter, eds. Wiley, New York.

Chomsky, N., and Schützenberger, M. P. (1962), "The Algebraic Theory of Context-Free Languages," P. Brafford and D. Hirschberg, eds. North Holland, Amsterdam.

Parikh, R. (1961), Language generating devices. R.L.E. Quart. Prog. Rept. No. 60, M.I.T., Cambridge, Mass.