Kybernetika

Jozef Gruska

On sets generated by context-free grammars

Kybernetika, Vol. 2 (1966), No. 6, (483)--493

Persistent URL: http://dml.cz/dmlcz/125712

KYBERNETIKA ČÍSLO 6, ROČNÍK 2/1966

On Sets Generated by Context-Free Grammars

JOZEF GRUSKA

The set E of strings is said to be definable (strongly definable) if there is a context-free grammar G such that E is the set of all terminal strings generated from the initial symbol (from all nonterminal symbols) of G. The properties of definable and strongly definable sets are investigated.

1. INTRODUCTION AND SUMMARY

Context-free grammars are a constructive means for the generation of the sets of strings in a given alphabet of so called terminal symbols. With a given context-free grammar it is possible to associate in a different way the set of terminal strings. Most frequently the case is investigated that an initial nonterminal symbol is given in a grammar G and the set $L(G)$ of all terminal strings generated from this initial symbol is considered to be the set generated by grammar G. The sets defined in such a way are called definable [6]. The properties of definable sets have been investigated in many papers [2, 6, 3].

In this paper also the set $L_s(G)$ is associated with a given grammar G; it is the set of all terminal strings which are generated by at least one nonterminal symbol of grammar G. The sets $L_s(G)$ are called strongly definable and it is said that grammar G strongly generates $L_s(G)$.

This paper is devoted to the study of definable and strongly definable sets. The technical results achieved in this paper are as follows:

In Section 2 two families of sets (of strings), the family of definable and strongly definable sets, are described and studied. The two families of sets are not identical although every strongly definable set is also definable. The relation of strongly definable sets to some classes of definable sets is investigated in Section 3. The closure properties of strongly definable sets are studied in Section 4. If A, B are definable sets and R is a regular set, then AB, $A - R$ are also definable. These results are shown not to be true if A, B are strongly definable. However, if A is strongly definable and R is finite then $A - R$ is also strongly definable (Theorem 3). One definable and one strongly definable set correspond to every grammar. It is shown to be undecidable whether they are identical (Section 5). In Section 6 some results about ambiguous definable sets are given. It is shown that there exists an ambiguous set which is strongly generated by unambiguous grammar. However it is not true for definable sets which are ambiguous of unbounded degree.

2. DEFINABLE AND STRONGLY DEFINABLE SETS

In this Section the basic concepts will be introduced. Especially, the classes of definable and strongly definable sets are defined and it is proved that they are not identical.

Let $\mathscr{A}$ be an alphabet. Its elements will be called symbols. A finite sequence of symbols will be called string over $\mathscr{A}$. ε will be the empty string. If x is a string then by $\lambda(x)$ we shall denote the length of x and by $x(i)$ the i-th symbol of x. If x, y are strings, $x = a_1 \dots a_n$, $y = b_1 \dots b_m$ where $a_1, \dots, a_n, b_1, \dots, b_m \in \mathscr{A}$, then xy denotes the string $a_1 \dots a_n b_1 \dots b_m$.

If A, B are two sets of strings over $\mathscr{A}$, then the set product AB of the sets A and B is defined by $AB = \{xy;\ x \in A,\ y \in B\}$. Moreover, let $A^\infty = \bigcup_{i=1}^{\infty} A^i$, $A^{\infty 0} = \bigcup_{i=0}^{\infty} A^i$ where $A^0 = \{\varepsilon\}$ and $A^{i+1} = A^i A$ for $i = 0, 1, \dots$

Definition 1. A context-free grammar G is a quadruple $\langle V_T, V_N, \mathscr{R}, S\rangle$ where V_T and V_N are non-empty, finite sets of terminal symbols and non-terminal symbols (or metasymbols), respectively; $V_T \cap V_N = \Lambda$ (the empty set); $\mathscr{R}$ is a finite set of (syntactical) rules $\alpha = A \to t$ where $A \in V_N$, $t \in (V_N \cup V_T)^\infty$ and $S \in V_N$ is the initial symbol of G.

(Occassionally, as when specifying a grammar for illustrative purposes, the given grammar G will be written as the set of rules $T_i \to t_i$, $i = 1, 2, \dots, n$ with T_1 as the initial symbol.)

A sequence $x_1, x_2, \dots, x_n$, $n > 1$ of strings from $(V_T \cup V_N)^\infty$ is said to be a derivation of x_n from x_1 if for every i, $1 \leqq i < n$ there are strings p_i, q_i and the rule $A_i \to t_i$ such taht $x_i = p_i A_i q_i$, $x_{i+1} = p_i t_i q_i$. If $x, y \in (V_T \cup V_N)^\infty$ then we write $x \underset{G}{\Rightarrow} y$ if and only if there is a derivation of y from x. (If there is no danger of misunderstanding, the symbol specifying the grammar will be deleted.)

For every $X \in V_N$ let $G(X) = \{x; X \Rightarrow x \in V_T^\infty\}$; $L(G) = G(S)$ and $L_s(G) = \bigcup_{X \in V_N} G(X)$. It is said that G generates the set $L(G)$ and strongly generates the set $L_s(G)$.

Definition 2. A set E is said to be definable (strongly definable) if there is a context-free grammar G such that $E = L(G)$ $(= L_s G)$*

Remark. The concept of strongly definable set is identical with the concept of the set generated by a context-free grammar within the meaning of paper [4] if one writes "e" not "$e \in E$" on page 47, [4], 12th last line. (It came to occur there inadvertently and alters the whole definition very disadvantageously although the results of paper [4] are correct.)

* Our concept of definable set is identical with the concept of definable set without ε used in paper [6].

Having specified two families of sets, namely, the family of definable sets and the family of strongly definable sets, it is natural to inquire as to whether or not the two families are equivalent. Since the sum of two definable sets is also definable (see [1]), one gets immediately:

Lemma 1. *Every strongly definable set is definable.*

The converse of this assertion is not true. To set that we shall prove

Lemma 2. *Let E be a set of strings and let x be a non-empty string such that the following conditions are satisfied for every integer n:*

(i) *there are strings P_n, Q_n (may be empty) such that $P_n x^n Q_n \in E$.*

(ii) *If y is a substring of x^n and $\lambda(y) \geqq \lambda(x)$, then for no string Q, $Qy \in E$, $yQ \in E$. Then E is not strongly definable.*

Proof. Suppose on the contrary that there is a grammar $G = \langle V_T, V_N, \mathscr{R}, S \rangle$ strongly generating E. Let $n > \max \{\lambda(t); U \to t \in \mathscr{R}\}$. According to (i) there are strings P_n and Q_n such that $P_n x^n Q_n \in E$. With respect to choice of n there is a rule $U \to u \in \mathscr{R}$ such that $u \Rightarrow P_n x^n Q_n$. But it is possible, with regard to choice of n, only if there are y and i such that y is a a substring of x^n, $\lambda(y) \geqq \lambda(x)$ and either $u(i) \Rightarrow$ $\Rightarrow Qy \in E$ or $u(i) \Rightarrow yQ \in E$ for some string Q. But this contradicts the condition (ii).

According to this lemma (taking $x = b$) the set $E_0 = \{ab^n a; n \geqq 1\}$ is not strongly definable although E_0 is obviously definable. This result and the result of Lemma 1 are summarized in the following theorem:

Theorem 1. *The family of strongly definable sets is a proper subset of the family of definable sets.*

Remark. Context-free grammars are used as a means for the definition of the syntax of artificial languages, especially, modern programming languages. For example the set of ALGOL's programs (if the limitations given in the non-formal part of Report [9] are not considered) is the definable set. In this connection there arises a question whether the set of texts of an artificial language is strongly definable. According to Lemma 2 (taking $x = a;$) the set of ALGOL's programs under discussion is not strongly definable. Similary (taking $y = a,$) the set of arithmetic expressions in ALGOL is not strongly definable. But if one considers the arithmetic expressions without functions and index variables, then we have a strongly definable set. This result indicates that strongly definable sets are not withouth interest in the study of artificial languages. For example the set of well-formed formulas of the propositional calculus (or of the predicate calculus), with a fixed finite number of variables is a strongly definable set.

An important practical problem, which arises in connection with the use of context-free grammars for the definition of the syntax of programming languages, is that of the construction of an algorithm for determining given an arbitrary grammar G and a string x, whether $x \in L(G)$. Such algorithms were constructed by many authors on the base of different principles. At some approaches to their construction (see, for example [2],) it is more easy to construct the algorithm for determining, given a grammar G and a string x, whether $x \in L_s(G)$. Hence if a set (of strings) M is strongly definable then, at least from the point of view the effective recognition of strings, it may be more convenient to construct a grammar G_1 which strongly generates M than a grammar G_2 which only generates M.

3. STRONGLY DEFINABLE SETS AND SPECIAL CLASSES OF DEFINABLE SETS

Often the special classes of definable sets, namely the classes of sets generated by finite state grammars,* linear grammars and metalinear grammars, respectively, are considered [3]. These sets will be called regular, linear and metalinear, respectively, and the relations between these classes and the class of strongly definable sets will be investigated.

The set E_0 from Section 2 is obviously a regular one and so there exists even a regular set which is not strongly definable. On the other hand, there exists a strongly definable set which is not regular. (Indeed, the set $E_1 = \{a^n b^n;\ n \geqq 1\}$ is not regular [6] but it is (strongly) generated by grammar G_1 with the rules $A \to aAb$, $A \to ab$.

A set E of strings will be called strongly regular if there is a finite state grammar G such that $L_s(G) = E$. According to previous results the family of strongly regular sets is a proper subset of the family of regular sets. The regular sets correspond to finite automata with a given initial state and strongly regular sets correspond to finite automata without an initial state. However, the family of strongly regular sets is not identical with the family of sets which are both strongly definable and regular. Indeed, the set $\{aa\}$ consisting of just one string is strongly definable and regular but not strongly regular.

As to the relation between the family of strongly definable sets and the families of linear and metalinear sets one gets the following: The set $E_1 = \{a^n b^n;\ n \geqq 1\}$ is strongly definable and linear but non regular [6]. Next, there is a strongly definable set which is metalinear but not linear. (The set $E_2 = \{a^n b^n a^m b^m;\ n + m > 0\}$ is (strongly) generated by grammar G_2 with the rules

$$S \to S_1 S_1\,, \quad S \to S_1\,, \quad S_1 \to a S_1 b\,, \quad S_1 \to ab$$

and hence E_2 is metalinear although it is, as can be easily seen, not linear). Finally, there is a strongly definable set which is not metalinear. By [3] the set (strongly) generated by grammar G_3 with the rules $S \to a$, $S \to bSS$ is not metalinear.

In all examples given above there were considered grammars G for which both sets $L(G)$ and $L_s(G)$ were of the same character. For example, both sets $L(G_1)$, $L_s(G_1)$ were linear but not regular. Is it so in general? The answer is in the negative. Indeed, consider grammar G_4 defined by the rules:

* Finite state grammars are defined as grammars all the rules of which are either of the form $A \to Ba$ or of the form $A \to a$ where A, B are non-terminal symbols and a is a terminal symbol.

$$S \to T\,,\quad S \to cU\,,\quad T \to aTb\,,\quad T \to ab\,,\quad U \to ab\,,\quad U \to aU\,,\quad U \to Ub$$

with the initial symbol S. Then the set $L(G_4)$ is linear but not regular and the set $L_s(G_4)$ is regular.

4. CLOSURE PROPERTIES

It is obvious that the sum of two strongly definable sets is again strongly definable. Moreover, if E is a strongly definable set, then the set E^∞ is also strongly definable. Indeed, let $G = \langle V_T, V_N, \mathscr{R}, S \rangle$ be a grammar such that $L_s(G) = E$. Let Z be a symbol such that $Z \notin V_T \cup V_N$. Denote G_1 the grammar defined by the quadruple $\langle V_T, V_N \cup \{Z\}, \mathscr{R} \cup \mathscr{R}_1, Z \rangle$, where $\mathscr{R}_1 = \{Z \to T;\ T \in V_N\} \cup \{Z \to ZZ\}$. It is readily seen that $L_s(G_1) = L_s(G)^\infty$. However, if A and B are strongly definable sets, then set product AB need not be strongly definable. Indeed, the set $E_0 = \{ab^na;\ n \geqq 1\}$ is not (see Section 2) strongly definable although $E_0 = A_0B_0$, where $A_0 = \{ab^n;\ n \geqq 1\}$ and $B_0 = \{a\}$ are strongly definable sets. (They are strongly generated by grammars G_1 with $\mathscr{R}_1 = \{A \to Ab, A \to ab\}$ and G_2 with $\mathscr{R}_2 = \{A \to a\}$, respectively).

Consider the grammars G_1 and G_2 defined as in (1) and (2), respectively:

(1) $$A \to aaAc\,,\quad A \to bAc\,,\quad A \to bc\,,$$

(2) $$A \to aAcc\,,\quad A \to aAb\,,\quad A \to ab\,.$$

The intersection of the sets (strongly) generated by grammars G_1 and G_2 is the set $\{a^{2n}b^nc^{2n};\ n \geqq 1\}$ which is not definable (see [3], Sec. 4.3). This example (along with the fact that the set union of two strongly definable sets and the iteration of a strongly definable set are strongly definable sets) establishes that

Theorem 2. *The family of strongly definable sets is closed under set union and iteration* but it is not closed under set product, intersection and complement.*

It is known [1] that if A is definable and B is regular, then $A - B$ is also definable. This result is not true for strongly definable sets. Indeed, if $A = \{ab^na;\ n \geqq 1\} \cup \{b\}^\infty$, $B = \{b\}^\infty$, then A is strongly definable, B is regular and $A - B$ is just the set E_0 from the Section 2 and hence $A - B$ is not strongly definable.

However, we have

Theorem 3. *If A is a strongly definable set and B is a finite set, then the set $A - B$ is strongly definable.*

The proof of this theorem will be omitted because it is rather cumbersome. It would be desirable to prove this theorem under weaker assumptions about B. However, this problem seems to be rather difficult because Theorem 3 is not valid, as the example given above shows, even for such simple infinite sets B as the set of all strings in an alphabet of just one symbol.

* If E is a set of strings then the set E^∞ will be called the iteration of E.

Finally, the influence of reduction and extension of gammars upon strongly definable sets associated with given grammars is investigated.*

If G_1 and G_2 are context-free grammars such that G_2 is a reduction of G_1, then $L_s(G_2) \subset L_s(G_1)$. Indeed, if $G_2 = \langle V_T, V_N, \mathscr{R}, S\rangle$, then $G_1 = \langle V_T, V_N \cup V'_N, \mathscr{R}', S'\rangle$ for some $V'_N, \mathscr{R}', S'$ and $G_1(T) = G_2(T)$ if $T \in V_N$. Therefore, $L_s(G_2) = \bigcup_{T \in V_N} G_2(T) = \bigcup_{T \in V_N} G_1(T) \subset \bigcup_{T \in V_N \cup V'_N} G_1(T) = L_s(G_1)$. Both cases: $L_s(G_1) \neq L_s(G_2)$ $(G_1 : S \to aU, U \to b;\ G_2 : S \to ab)$ and $L_s(G_1) = L_s(G_2)$, $(G_1 : S \to aUc,\ S \to b,\ U \to b;\ G_2 : S \to abc,\ S \to b)$ are possible. The same is true for the case that G_1 is an extension of G_2.

5. UNDECIDABLE PROPERTIES

It is easily seen that for a given grammar G the sets $L(G)$ and $L_s(G)$ have not to be equal. However, as remarked in the Introduction, there is no algorithm for determining whether equality holds (Theorem 5). Moreover, it shall be proved that it is true even when only "well-formed" grammars, i.e. grammars which have no needless rule are considered.

Definition 3. A grammar $G = \langle V_T, V_N, \mathscr{R}, S\rangle$ is said to be well-formed if for every $a \in V_T \cup V_N$ there are strings x, y and integer i such that $S \Rightarrow x \underset{=}{\Rightarrow} y \in V_T^\infty$, $x(i) = a$.

Remark. For a given grammar G it is easy to construct a new grammar G_1 (by omitting those rules which are not used in generating the set $L(G)$), such that G_1 is well-formed and $L(G) = L(G_1)$. However, in general the sets $L_s(G)$ and $L_s(G_1)$ need not be equal.

Theorem 4. *There is no alogorithm for determining given an arbitrary well -formed grammar G, whether $L(G) = L_s(G)$.*

Proof. Let $G = \langle V_T, V_N, \mathscr{R}, S\rangle$ be a well-formed grammar such that there are two different symbols a_0, a_1 in V_T. To grammar G we can construct well-formed grammar $G^1 = \langle V_T, V_N^1, \mathscr{R}^1, S_1\rangle$ such that $V_N^1 \cap V_N = \Lambda$, $L(G^1) = V_T^\infty$. Now let $Z, Z_1 \notin (V_T \cup V_N \cup V_N^1)$ be two different symbols. Denote for $i = 0, 1$:

$$G^{a_i} = \langle V_T, V_N \cup V_N^1 \cup \{Z, Z_1\}, \mathscr{R} \cup \mathscr{R}^1 \cup \{Z \to S, Z \to Z_1, Z_1 \to a_i S_1\}, Z\rangle .$$

G^{a_i} are well-formed grammars. There is

Lemma 3. $L(G) = V_T^\infty$ *if and only if*

(i) $$L(G) = L_s(G) \quad \text{and} \quad L(G^{a_i}) = L_s(G^{a_i}) \quad \text{for} \quad i = 0, 1 .$$

Proof of the Lemma. First let (i) hold and suppose that there is a $x \in V_T^\infty - L(G)$.

* For the concepts "reduction" and "extension" of context-free grammars see paper [5].

Let $0 \leqq i \leqq 1$ be such that $x(1) \neq a_i$. Consider grammar G^{a_i}. Since $L(G^{a_i}) = L_s(G^{a_i})$ and $G^{a_i}(S_1) = V_T^\infty$ one has $L(G^{a_i}) = V_T^\infty$ and hence $x \in L(G^{a_i})$. Since $V_N \cap V_N^1 = \Lambda$ and $G^{a_i}(Z_1)$ contains all strings from V_T^∞ with the exception of the strings beginning with a_i, one has $x \in G^{a_i}(S) = G(S) = L(G)$ which is a contradiction and hence $L(G) = V_T^\infty$. On the other hand, if $L(G) = V_T^\infty$, then obviously (i) holds and this completes the proof of the Lemma.

Now we can proceed in proving the Theorem. Suppose that there is an algorithm for determining, given an arbitrary well-formed grammar G, whether $L(G) = L_s(G)$. Then, with regard to Lemma 3, there is an algorithm for determining, given an arbitrary well-formed grammar $G = \langle V_T, V_N, \mathscr{R}, S\rangle$ such that V_T consists of at least two symbols, whether $L(G) = V_T^\infty$. However, it is known [3] that there is no algorithm for determining given a grammar $G = \langle V_T, V_N, \mathscr{R}, S\rangle$, whether $L(G) = V_T$. But the proof is valid even if one considers only well-formed grammars which have at least two terminal symbols. Hence the assumption that the assertion of the Theorem does not hold yields a contradiction and the Theorem is proved.

From this one derives the following results immediately.

Theorem 5. *There is no algorithm for determining, given an arbitrary grammar G, whether $L(G) = L_s(G)$.*

Corollary 6. *There is no algorithm for determining, given a grammar $G = \langle V_T, V_N, \mathscr{R}, S\rangle$ and $S_1 \in V_N - \{S\}$ such that $S \Rightarrow x$, $x(i) = S_1$ for suitable x and i, whether $G(S_1) \subset G(S)$.*

6. AMBIGUITY

Definition 3. A definable set E is called ambiguous of the degree n where n is an integer or ∞ (in the last case it will also be said that E is ambiguous of unbounded degree) if the following conditions are satisfied:

(i) If a grammar G generates E and $n \neq \infty$ ($n = \infty$) then, (for every integer m) there is a string $x \in E$ which is generated in at least n (m if $n = \infty$) essentially different ways, that is, x has n distinct structural descriptions (see for example [3], p. 367).

(ii) If $n \neq \infty$, then there is a grammar G generating E and such that every string $x \in E$ is generated in at most n essentially different ways.

A definable set is called unambiguous if it is ambiguous of the degree 1 and it is called ambiguous if it is not unambiguous. (If a set E is ambiguous of the degree n we shall also say that E is n-ambiguous).

Consider grammars G_0 and G_1 defined as in (3) to (6) and (4) to (7), respectively:

(3) $\qquad S \to S_1S_2\,, \qquad S \to S_2S_3\,,$

(4) $\qquad S_1 \to aS_1b\,, \qquad S_1 \to ab\,,$

(5) $S_2 \to S_2a\,, \quad S_2 \to a\,,$

(6) $S_3 \to bS_3a\,, \quad S_3 \to ba\,,$

(7) $S \to S_1S_2\,, \quad S_4 \to S_2S_3\,.$

The set

$$M = \{a^nb^ma^p;\ n = m \quad \text{or} \quad m = p\} \cup \{a^nb^n;\ n \geqq 1\} \cup \{b^na^n;\ n \geqq 1\} \cup \{a\}^\infty$$

is strongly generated by ambiguous grammar G_0 and also by unambiguous grammar G_1. The set M is ambiguous (of the degree 2). Indeed, according to [7], if E is an unambiguous definable set and B is a regular set, then $E - B$ is unambiguous, too. The set $N = \{a\}^\infty \{b\}^\infty \cup \{b\}^\infty \{a\}^\infty \cup \{a\}^\infty$ is regular and the set $M - N = \{a^nb^ma^p;\ n = m \text{ or } m = p\} = P$ is ambiguous [3]; hence M is also ambiguous (of the degree 2 as can be easily seen). Consequently, one gets the result:

Theorem 7. *There is an ambiguous (of the degree 2), definable set which is strongly generated by an unambiguous grammar.*

Definition 4. A grammar G is said to be completely unambiguous if G is unambiguous and if for every $x \in L_s(G)$ there is just one metasymbol X such that $X \Rightarrow x$.

The conditions for a grammar to be completely unambiguous seem to be very strong. But one gets:

Theorem 8. *There is an ambiguous (of the degree 2) definable set which is strongly generated by a completely unambiguous grammar.*

Proof. The grammar G_1 given above is completely unambiguous; G_1 strongly generates M and M is ambiguous.

Moreover, there is

Theorem 9. *There is no algorithm for determining, given an arbitrary grammar G, whether or not G is completely unambiguous.*

Proof. If such algorithm exists, then there is an algorithm to decide whether an arbitrary grammar with one metasymbol is unambiguous, contrary see in [8].

As to the sets strongly generated by unambiguous grammars one gets:

Theorem 10. *If a set E is strongly generated by an unambiguous grammar, then E is not unboundedly ambiguous.*

Proof. Let unambiguous grammar $G = \langle V_T, V_N, \mathscr{R}, S\rangle$ strongly generate E. Let $Z \notin (V_N \cup V_T)$ be a symbol. Then the grammar $G' = \langle V_T, V_N \cup \{Z\}, \mathscr{R} \cup \{Z \to U, U \in V_N\}, Z\rangle$ generates E and obviously every string $x \in E$ is generated in at most n essentially different ways where n is the cardinality of the set V_N.

Remark. Theorem 11 given below indicates that presumably it is not possible to prove a stronger result.

Presumably

(H 1) the set $P_n = P^n \cup \{a^nb^n;\ n \geqq 1\} \cup \{b^na^n;\ n \geqq 1\} \cup \{a\}^\infty$ is $(2n)$-ambiguous.

However, the set P_n is strongly generated by an unambiguous grammar. Only the unambiguous grammar G_2 strongly generating P_2 is given. But from the construction of G_2 it is obvious how to construct grammar G_n strongly generating P_n. The grammar G_2 is given by the rules.

$$S_1 \to XZXZ\,, \quad S_2 \to XZY\,, \quad S_3 \to ZYZY\,, \quad S_4 \to ZYXZ\,,$$

$$X \to aXb\,, \qquad X \to ab\,,$$

$$Y \to bYa\,, \qquad Y \to bA\,,$$

$$Z \to Za\,, \qquad Z \to a\,.$$

Hence:

Theorem 11. *If* (H 1) *holds, then for every integer n there are* $2n$*-ambiguous sets strongly generated by an unambiguous grammar.*

A natural question is whether a strongly definable set exists which is neither generated nor strongly generated by an unambiguous grammar. According to Theorem 10, in order to prove that such set exists, it suffices to prove that there is a strongly definable set which is ambiguous of unbounded degree. The set

$$P_\infty = P^\infty \cup \{a^n b^n;\ n \geqq 1\} \cup \{b^n a^n;\ n \geqq 1\} \cup \{a\}^\infty$$

is strongly generated by the grammar G_∞ defined by the rules

$$S \to SS\,, \qquad S \to S_1 S_3\,, \quad S \to S_3 S_2\,,$$

$$S_1 \to aS_1b\,, \quad S_1 \to ab\,,$$

$$S_2 \to bS_2a\,, \quad S_2 \to ba\,,$$

$$S_3 \to S_3a\,, \quad S_3 \to a\,.$$

Presumably:

(H 2) The set P_∞ is ambiguous of unbounded degree.

Consequently, one gets the result:

Theorem 12. *If* (H 2) *holds, then there is a strongly definable set which is neither generated nor strongly generated by an unambiguous grammar.*

A question arises whether for well-formed grammars G both sets $L(G)$ and $L_s(G)$ are ambiguous of the same degree. For the grammar G_0 defined as in (3) to (6) both sets $L(G_0)$ and $L_s(G_0)$ are ambiguous (of the degree 2) but the following examples show that it is not true in general.

Example 1. Let the grammar G_3 be defined by the rules (3) to (6) and the following rules:

$$S \to S_4 , \quad S_4 \to cS_5c , \quad S_5 \to S_6S_2 , \quad S_5 \to S_2S_7 ,$$

$$S_6 \to aS_6 , \quad S_6 \to S_6b , \quad S_6 \to ab ,$$

$$S_7 \to bS_7 , \quad S_7 \to S_7a , \quad S_7 \to ba .$$

Then the set $L(G_3) = P \cup \{ca^ib^ja^kc;\ i, j, k \geqq 1\}$ is ambiguous and $L_s(G_3) = \{ca^ib^ja^kc;\ i, j, k \geqq 1\} \cup \{a^lb^ma^n;\ n + m + p > 0\}$ is regular and thus unambiguous.

Example 2. Let the grammar G_4 be defined by the rules:

$$S \to S_1S_2 , \quad S \to cS_4 , \quad S_1 \to aS_1b , \quad S_1 \to ab , \quad S_2 \to aS_2 ,$$

$$S_2 \to a , \quad S_4 \to S_2S_3 , \quad S_3 \to bS_3a , \quad S_3 \to ba .$$

Then $L(G_4) = \{a^nb^na^p;\ n, p \geqq 1\} \cup \{ca^nb^pa^p;\ n, p \geqq 1\}$ is unambiguous but $L_s(G_4) = M \cup \{ca^nb^pa^p;\ n, p \geqq 1\}$ is ambiguous.

(Received August 15th, 1965)

REFERENCES

[1] Bar - Hillel, Perlis and Shamir: On formal properties of simple phrase structure grammars. Tech. Rep. No. 4 (July 1960). Applied Logic Branch, The Hebrew University of Jerusalem.
[2] N. Chomsky: On certain Properties of Grammars. Information and Control *2* (1959), 137 až 167.
[3] N. Chomsky: Formal Properties of Grammars. Handbook of Mathematical Psychology Vol. 2, Willey, New York, 1963.
[4] K. Čulík: Semantics and Translation of Grammars and ALGOL-like languages. Kybernetika *1* (1965), 47–49.
[5] K. Čulík: On equivalent and similar grammars of ALGOL-like languages. Comm. Math. Univ. Carol. *5* (1964), 57–59.
[6] S. Ginsburg and H. Rice: Two Families of Languages Related to ALGOL. JACM *9* (1962), 350–371.
[7] S. Ginsburg: The theory of context-free languages (In print).
[8] Sheita A. Greibach: The Undecidability of the Ambiguity Problem for Minimal Linear Grammars. Information and Control *6* (1963), 119–125.
[9] Revised Report on ALGOL 60. Comm. ACM *5* (1962), 299–315.

O množinách generovaných bezkontextovou gramatikou

JOZEF GRUSKA

Množina reťazcov E sa nazýva definovateľnou (silne definovateľnou) ak existuje bezkontextová gramatika G taká, že E je množinou všetkých terminálnych teťazcov odvodených z daného neterminálneho symbola (zo všetkých neterminálnych symbolov) gramatiky G.

V práci sa vyšetrujú silne definovateľné množiny; ich vzťah k definovateľným množinám (tj. k bezkontextovým jazykom) a ich špeciálnym triedam, uzáverové vlastnosti, niektoré rozhodovacie problémy a otázky viacznačnosti.

RNDr. Jozef Gruska CSc., Ústav technickej kybernetiky SAV, Bratislava, Dúbravská cesta