



Soviet-era science, translated into English

A. Ya. Dikovsky

1970

SovietRxiv

View the original and related papers at <https://sovietrxiv.org/items/ru-197001.90479>

Source: Math-Net.Ru and CyberLeninka. Machine translation. Verify with the original.

Abstract

Full Text

A. Ya. Dikovskiy

LANGUAGES OF BOUNDED ACTIVE CAPACITY

(Presented by Academician P. S. Novikov, 15 XII 1969)

In this note we investigate the structural properties of languages generated by grammars with certain restrictions on the “working capacity” of a derivation.

We shall consider generative grammars with rules of the form $\omega_1\varphi\omega_2 \rightarrow \omega_1\psi\omega_2$, where ω_1, ω_2 are strings in the basic alphabet, and φ is a nonempty string whose leftmost and rightmost symbols are auxiliary ones (as a particular case, this includes the generative grammars often considered in the literature with rules of the form $\varphi \rightarrow \psi$, where φ is a nonempty string of auxiliary symbols).

Let $\Gamma = (V, V_1, I, R)$ be some grammar of the class under consideration. By the active capacity of a string φ in the alphabet $V \cup V_1$ (notation $|\varphi|_1$) we shall mean the number of occurrences in φ of auxiliary symbols, and by the active capacity of a derivation $(\varphi_1, \varphi_2, \dots, \varphi_n)$ in Γ the number

$$s = \max_{1 \leq i \leq n} \{|\varphi_i|_1\}.$$

By the active capacity of a string $x \in L(\Gamma)$ (notation $s_{\text{inf}}(x)$) we shall mean the least active capacity of a complete derivation of x in Γ (i.e., of a derivation of x from the initial symbol I). If there exists a number k such that for every string $x \in L(\Gamma)$, $s_{\text{inf}}(x)$ does not exceed k , then we shall say that Γ is a grammar of bounded active capacity; moreover, the least of all such numbers k will be called the active capacity of Γ and denoted $s_{\text{inf}}(\Gamma)$. We shall also be interested in the case where there exists a number l that bounds from above the active capacity of any complete derivation in Γ ; then we shall say that Γ is a grammar of uniformly bounded active capacity, and the least of all such numbers l will be called the limiting active capacity of Γ and denoted $s_{\text{sup}}(\Gamma)$.

Theorem 1. *For every grammar Γ of bounded (uniformly bounded) active capacity one can construct an equivalent cs-grammar Γ' of bounded (uniformly bounded) active capacity, and*

$$s_{\text{inf}}(\Gamma') \leq s_{\text{inf}}(\Gamma)$$

(respectively

$$s_{\text{sup}}(\Gamma') \leq s_{\text{sup}}(\Gamma)$$

).

Theorem 1 allows us, without loss of generality, to restrict ourselves to the class of cs-grammars of bounded active capacity.

A convenient tool for investigating cs-grammars of bounded active capacity is the following measure of the complexity of a derivation tree—we shall call it density and denote it by μ .

Let $\Gamma = (V, V_1, I, R)$ be an arbitrary cs-grammar and let γ be the tree of a complete derivation in Γ of a string x (notation $\gamma = (I \leq x)$).

1. If α is a terminal vertex of γ , then $\mu(\alpha) = 0$.
2. Suppose arcs from α lead to the vertices $\alpha_1, \alpha_2, \dots, \alpha_r$; then

$$\mu(\alpha) = 1 + \max_{1 \leq l \leq r} \{\mu(\alpha_l)\},$$

if there exist i and j , $1 \leq i \neq j \leq r$, such that

$$\mu(\alpha_i) = \mu(\alpha_j) = \max_{1 \leq l \leq r} \{\mu(\alpha_l)\},$$

and

$$\mu(\alpha) = \max_{1 \leq l \leq r} \{\mu(\alpha_l)\}$$

otherwise. The density $\mu(\gamma)$ of the derivation tree γ is the density $\mu(\beta)$ of its root β .

The density $\mu(\gamma)$ of the complete derivation tree γ turns out to be very simply related to its active capacity $s_{\text{inf}}(\gamma)$ (i.e., to the least active capacity of a derivation produced by the tree γ).

Theorem 2. For every cs-grammar Γ , the maximum length of the right-hand sides of whose rules is equal to d , and for every complete derivation tree γ in Γ ,

$$\mu(\gamma) \leq s_{\text{inf}}(\gamma) \leq d[\mu(\gamma) + 1].$$

Theorem 2 gives a simple criterion for bounded active capacity of a cs-grammar Γ , consisting in the fact that there is a number c such that, for every string $x \in L(\Gamma)$, there exists a derivation tree

$$\gamma = \left(I \stackrel{\Gamma}{\Leftarrow} x \right)$$

with density $\mu(\gamma) \leq c$.

E. D. Stotskii ⁽¹⁾ posed the question of the relationship between the class of all cs-languages and the class of languages of bounded active capacity, i.e., languages generated by cs-grammars of bounded active capacity. This question is answered by

Theorem 3. No cs-grammar of bounded active capacity generates the language L_0 of all binary bracket sequences (i.e., the language: a) containing the sequence

(), b) containing the sequence ($z'z''$) for any sequences z' and z'' belonging to it, and c) containing no other sequences).

An interesting feature of languages of bounded active capacity is that they are “constructed” from linear languages by means of the operation of substitution (a mapping h of a language over the alphabet $V = \{a_1, \dots, a_n\}$ into some other language is called a substitution if, for every i , $1 \leq i \leq n$, $h(a_i)$ is some language and

$$h(a_{i_1} \dots a_{i_s}) = h(a_{i_1}) \dots h(a_{i_s}).$$

More precisely, the following holds.

Theorem 4. The class of languages of bounded active capacity coincides with the smallest class of languages containing all linear languages and closed under the operation of substitution.

In the proof of this theorem, Theorem 2 is used in an essential way.

The further results concern languages generated by cs-grammars of uniformly bounded active capacity (languages of uniformly bounded active capacity).

The question of the relationship between the classes of languages of bounded and uniformly bounded active capacity is resolved as follows.

Theorem 5. For any number $k = 1, 2, \dots$, the language

$$\{a^n b^n \mid n = 1, 2, \dots\}^k$$

is generated by a suitable cs-grammar Γ with limiting active capacity k , but is not generated by any cs-grammar Γ' with limiting active capacity less than k .

From this theorem follows the obvious

Corollary. The language

$$L_1 = \bigcup_{j=1}^{\infty} \{a^n b^n \mid n = 1, 2, \dots\}^j$$

is not generated by any cs-grammar of uniformly bounded active capacity.

It remains to note that the language L_1 is generated by a cs-grammar with active capacity two.

Languages of uniformly bounded active capacity have a structural characterization similar to that obtained above for the class of all languages of bounded active capacity. As it turns out, the languages of this class are constructed from linear languages by means of the operations of union, product, and substitution into a centered linear language (a linear language L over the alphabet $V = V' \cup \{C\}$, $C \notin V'$, is centered if $L \subseteq V'^* c V'^*$, c is called the center of the language L ; h is a substitution into a centered linear language L if h maps the center L into some language, and on the remaining symbols of the alphabet is identical). In exact form this statement is as follows:

Theorem 6. The class of languages of uniformly bounded active capacity coincides with the smallest class of languages containing all linear languages and closed under the operations of union, product, and substitution into a centered linear language.

In conclusion, we give one more result, somewhat outside the form of our presentation, but also providing a certain characterization of the class of languages of uniformly bounded active capacity.

Let $\Gamma = (V, V_1, I, R)$ be a generative grammar of the type indicated above. Associate with each rule in R some symbol r in such a way that distinct rules correspond to distinct labels; denote the set of all labels by \bar{R} . Then to every complete derivation D in Γ there is naturally assigned a string over the alphabet \bar{R} —the characteristic of D , and to the grammar Γ there corresponds the set of all characteristics of complete derivations in Γ —the superstructure of Γ .

Theorem 7. *A language L is a language of uniformly bounded active capacity if and only if it is generated by some grammar whose superstructure is a regular language.*

This theorem, in a somewhat weaker form, is also proved in (2).

Institute of Mathematics
Siberian Branch of the Academy of Sciences of the USSR
Novosibirsk

Received
11 XII 1969

References

¹ E. D. Stotskii, *Scientific and Technical Information*, ser. 2, 5 (1969). ² J. Friant, Rapport MA-102, CETADOL, Univ. de Montréal (1968).

Note: Figure translations are in progress. See original paper for figures.

Source: Math-Net.Ru and CyberLeninka. Machine translation. Verify with the original.