CONTEXTUAL GRAMMARS

- Solomon Marcus -

Institutul de Matematica
Str, Mihai Eminescu, 47
Bucharest 9, ROMANIA

In the following, we shall introduce a type of generative grammars, called contextual grammars. They are not comparable with regular grammars. But every language generated by a contextual grammar is a context-free language. Generalized contextual grammars are introduced, which may generate non-context-free languages.

Let V be a finite non-void set ; V is called a vocabulary. Every finite sequence of elements in V is said to be a string on V. Given a string $x = a_1 a_2 \ldots a_n$, the number n is called the length of x. The string of length zero is called the null-string and is denoted by ω . Any set of strings on V is called a language on V. The set of all strings on V (the null-string inclusively) is called the universal language on V. By a^n we denote the string $a \ldots a$, where a is iterated n times.

Any ordered pair $\langle u,v \rangle$ of strings on V is said to be a context on V. The - string x is admitted by the context $\langle u,v \rangle$ with respect to the language L if $uxv \in L$.

Let L_1 be a finite set of strings on the vocabulary V and let $\mathcal{C}$ be a finite set of contexts on V. The triple

$$(V, L_1, \mathcal{C}) \qquad (1)$$

is said to be a contextual grammar ; V is the vocabulary of the grammar, L_1 is the base of the grammar and $\mathcal{C}$ is the contextual component of the grammar.

Let us denote by G the contextual grammar defined by (1). Consider the smallest language L on V, fulfilling the following two conditions

(α) $L_1 \subseteq L$;

(β) if $x \in L$ and $\langle u,v \rangle \in \mathcal{C}$, then $uxv \in L$.

The language L is said to be the language generated by the contextual grammar G. This means that the language generated by G is the intersection of all languages L fulfilling the conditions (α) and (β).

A language L is said to be a contextual language if there exists a contextual grammar G which generates L.

Proposition 1. Every finite language is a contextual language.

Proof. Let V be a vocabulary and let L_1 be a finite language on V. It is obvious that the contextual grammar $(V, L_1, 0)$, where 0 denotes the void set of contexts, generates the language L_1. The same language may be generated by means of the contextual grammar $(V, L_1, \mathcal{U})$, where $\mathcal{U}$ is formed by the null-context only.

Two contextual grammars are called equivalent if they generate the same language. The grammars $(V, L_1, 0)$ and $(V, L_1, \mathcal{U})$ are equivalent, since they both generate the language L_1.

The converse of Proposition 1 is not true. Indeed, we have

Proposition 2. The universal language is a contextual language.

Proof. Let $V = \{a_1, a_2, \ldots, a_n\}$. Denote by L the universal language on V. Let us put $L_1 = \{\omega\}$ and $\mathcal{C} = \{ \langle \omega, a_1 \rangle$,

$\langle\omega, a_2\rangle, \ldots, \langle\omega, a_n\rangle\}$. It is easy to see that the grammar $(V, L_1, \mathcal{C})$ generates the universal language on V.

Remarks. If we put, in the proof of Proposition 2, $L_1 = V$ instead of $L_1 = \{\omega\}$, then the grammar $(V, L_1, \mathcal{C})$ does not generate the universal language on V, since the language it generates does not contain the null-string.

In order to illustrate the activity of the grammar $(V, L_1, \mathcal{C})$ defined in the proof of Proposition 2, let us consider the particular case when the vocabulary is formed by two elements only : $V = \{a, b\}$. The general form of a string x on V is $x = a^{i_1}b^{j_1}a^{i_2}b^{j_2}\ldots a^{i_p}b^{j_p}$, where $i_1, j_1, i_2, j_2, \ldots, i_p, j_p$ are arbitrary non-negative integers. In order to generate the string x, we start with the null-string ω and we apply i_1 times the context $\langle\omega, a\rangle$. The result of this operation is the string a^{i_1}, to which we apply j_1 times the context $\langle\omega, b\rangle$ and obtain the string $a^{i_1}b^{j_1}$. Now we apply i_2 times the context $\langle\omega, a\rangle$, then j_2 times the context $\langle\omega, b\rangle$ and we continue so alternatively. When, after $2p-2$ steps, we have obtained the string $y = a^{i_1}b^{j_1}a^{i_2}b^{j_2}\ldots a^{i_{p-1}}b^{j_{p-1}}$, it is enough to apply i_p times the context $\langle\omega, a\rangle$ and, to the string so obtained, j_p times the context $\langle\omega, b\rangle$, in order to generate completely the string x.

Haskell Curry considered the language $L = \{ab^n\}$ ($n = 1, 2, \ldots$) as a model of the set of natural numbers [5]. We call L the language of Curry.

Proposition 3. The language of Curry is a contextual language.

Proof. The considered language is generated by the grammar $(V, L_1, \mathcal{C})$, where $V = \{a.b\}$, $L_1 = \{a\}$ and $\mathcal{C} = \{\langle\omega, b\rangle\}$.

We recall that a language is said to be regular if it may be generated by means of a finite automaton (or, equivalently, by means of a finite state grammar in the sense of Chomsky).

Proposition 4. There exists a contextual language which is not regular.

Proof. Let us consider the language $L = \{a^n b^n\}$ $(n=1,2,\ldots)$ If we put $V = \{a,b\}$, $L_1 = \{ab\}$ and $\mathcal{C} = \{\langle a.b\rangle\}$, then it is easy to see that L is generated by the contextual grammar $(V, L_1, \mathcal{C})$. On the other hand, L is not a regular language. This fact was asserted by Chomsky in [3] and [4], but the proof he gives is wrong. A correct proof of this assertion and a discussion of Chomsky's proof were given in [8], and [9].

Propositions 2, 3 and 4 show that there are many infinite languages which are contextual. This fact may be explained by means of

Proposition 5. If the set L_1 is non-void and if the set $\mathcal{C}$ contains at least one non-null context, then the contextual grammar $(V.L_1, \mathcal{C})$ generates an infinite language.

Proof. Since L_1 is non-void, we may find a string x belonging to L_1. Since $\mathcal{C}$ contains at least one non-null context, let $\langle u, v\rangle$ be a non-null context belonging to $\mathcal{C}$. From these assumptions, we infer that the strings

$$uxv\ ,\ u^2xv^2, \ldots, u^nxv^n, \ldots$$

are mutually distinct and belong all to the language generated by the grammar $(V, L_1, \mathcal{C})$. Thus, this language is infinite.

The converse of Proposition 5 is true. Indeed, we have

Proposition 6. If the contextual grammar $(V, L_1, \mathscr{C})$ generates an infinite language, then L_1 is non-void, whereas $\mathscr{C}$ contains a non-null context.

Proof. Let L be the language generated by $(V, L_1, \mathscr{C})$. If L_1 is void, L is void too, hence it cannot be infinite. If $\mathscr{C}$ contains no non-null context, we have $L = L_1$. But L_1 is in any case finite ; thus, L is finite, in contradiction with the hypothesis.

Since there are contextual language which are not regular (see Proposition 4 above), it would be interesting to establish whether all contextual languages are context-free languages. The answer is affirmative :

Proposition 7. Every contextual language is a context-free language.

Proof. Let L be a contextual language. If L is finite, it is a regular language. But it is well known that every regular language is a context-free language. Therefore, L is a context-free language. Nowe let us suppose that L is infinite. Denote by $G = (V, L_1, \mathscr{C})$ a contextual grammar which generates the language L. In view of Proposition 6, L_1 is non-void, whereas there exists an integer i, $1 \leq i \leq p$, such that the context $\langle u_i, v_i \rangle$ is non-null, i.e. at least one of the equalities $u_i = \omega$, $v_i = \omega$ is false. Let us make a choice and suppose that $u_i \neq \omega$. Let $L_1 = \{x_1, x_2, \ldots, x_n\}$ and $\mathscr{C} = \{\langle u_1, v_1 \rangle, \langle u_2, v_2 \rangle, \ldots, \langle u_p, v_p \rangle\}$. We define a context-free grammar Γ as follows. The terminal vocabulary of Γ is V. The non-terminal vocabulary of Γ contains one element only - denoted by S - which is, of course, the axiom of the grammar Γ. The terminal rules of Γ are

$$S \to x_1 ,$$
$$S \to x_2 ,$$
$$\cdots\cdots$$
$$S \to x_n$$

whereas the non-terminal rules are

$$S \to u_1 S v_1 ,$$
$$S \to u_2 S v_2 ,$$
$$\cdots\cdots\cdots$$
$$S \to u_p S v_p .$$

It is obvious that the number of terminal rules is equal to the number of strings in L_1 , whereas the number of non-terminal rules is precisely the number of contexts in $\mathcal{C}$. Among the non-terminal rules, there is one at least which is non-trivial : it is the rule $S \to u_i S v_i$, where $u_i \neq \omega$.

It is not difficult to prove that the grammar Γ generates the given language L. Indeed, the general form of a string in L is

$$u_{i_1}^{j_1} u_{i_2}^{j_2} \cdots u_{i_p}^{j_p} \; y \; v_{i_p}^{j_p} \cdots v_{i_2}^{j_2} v_{i_1}^{j_1} \qquad (2)$$

where $y \in V$ and

$$\langle u_{i_s} , v_{i_s} \rangle \in \mathcal{C} \quad \text{for } s = 1,2,\ldots,p.$$

In order to generate the considered string we begin by applying j_1 times the rule

$$S \to u_{i_1} S v_{i_1} .$$

In this way, we obtain the expression

$$u_{i_1}^{j_1} S \, v_{i_1}^{j_1} .$$

The next step consists in applying j_2 times the rule

$$S \rightarrow u_{i_2} S v_{i_2} ,$$

which yields the expression

$$u_{i_1}^{j_1} u_{i_2}^{j_2} S v_{i_2}^{j_2} v_{i_1}^{j_1} .$$

Continuing in this way, we arrive, after p-1 steps, to the expression

$$u_{i_1}^{j_1} u_{i_2}^{j_2} \ldots u_{i_{p-1}}^{j_{p-1}} S v_{i_{p-1}}^{j_{p-1}} \ldots v_{i_2}^{j_2} v_{i_1}^{j_1} .$$

We now apply j_p times the rule

$$S \rightarrow u_{i_p} S v_{i_p}$$

and thus we obtain the expression

$$u_{i_1}^{j_1} u_{i_2}^{j_2} \ldots u_{i_{p-1}}^{j_{p-1}} u_{i_p}^{j_p} S v_{i_p}^{j_p} v_{i_{p-1}}^{j_{p-1}} \ldots v_{i_2}^{j_2} v_{i_1}^{j_1}$$

where, by applying the terminal rule

$$S \longrightarrow y,$$

the considered string is completely generated. Thus, we have proved that L is contained in the language generated by Γ.

Conversely, let z be a string generated by Γ. The general form of this generation involves several consecutive applications of non-terminal rules (the number of these applications may be eventually equal to zero) followed by one and only one application of a terminal rule. It is easy to see that the result of this generation is always a string of the form (2). Thus we have proved that the language generated by Γ is contained in L. In view of the preceding considerations, L is precisely the language generated by Γ.

Proposition 7 easily permits to obtain simple examples of languages which are not contextual languages. For instance, the language of Kleene $\{a^{n^2}\}$ $(n=1,2,\ldots)$, the first example of an infinite language which is not regular, is a very simple example of non-contextual language. It is enough to remark that the sequence $\{n^2\}$ $(n = 1,2,\ldots)$ contains no subsequence which is an infinite arithmetic progression (We have $(n+1)^2-n^2=2n+1$ and $\lim\limits_{n\to\infty}(2n+1)=\infty$, therefore for every subsequence of $\{n^2\}$ the difference of two consecutive terms has the limit equal to $+\infty$ when $n\to\infty$). But a result of [1] asserts, among others, that given an infinite context-free language L, the set of integers which represent the lengths of the strings in L contains an infinite arithmetic progression. It follows that the language of Kleene is not context-free and, in view of Proposition 7, it is not a contextual language. The same fact follows from theorem 3.1.2 of [6], p. 86.

A natural question now arrises : Do there exist non-contextual languages among context-free languages ? The affirmative answer follows from the following remark :

The converse of Proposition 7 is not true. Indeed, we have

Proposition 8. There exists a context-free language which is not a contextual language.

Proof. Let $V = \{a,b\}$. In view of a theorem of Gruška [7] there exists, for every positive integer n, a context-free language L_n on V, such that every context-free grammar of L_n contains at least n non-terminal symbols. But, as we can see in the proof of Proposition 7, every contextual language may be generated with a context-free grammar containing only one non-terminal symbol. Therefore, if $n \geq 2$, L_n is not a contextual language.

Proposition 8 suggests the natural question whether there exist regular languages which are not contextual languages. The answer is affirmative :

Proposition 9. There exists a regular language which is not a contextual language.

Proof. Let us consider the language $L = \{ab^m \frown c \frown a \frown b^n\}$ $(m,n=1,2,\ldots)$, which was used by H.B.Curry [5], in order to describe the set of mathematical (true or not) propositions. This language is regular, since it can be generated by the rules $S \to Ab$, $A \to Ab$, $A \to Ba$, $B \to Cc$, $C \to Cb$, $C \to Db$, $D \to a$. We shall show that L is not a contextual language. Indeed, let us admit that the contrary holds and let $G = \langle V, L_1, \mathcal{C} \rangle$ be a contextual grammar of L. Here, the general form of a string in L is

$$u_n^{p_n} \ldots u_2^{p_2} u_1^{p_1} \, x \, v_1^{p_1} v_2^{p_2} \ldots v_n^{p_n} \qquad (3)$$

where $x \in L_1$, whereas $\langle u_i, v_i \rangle \in \mathcal{C}$ $(i = 1,2,\ldots,n)$ and $p_1, p_2, \ldots, p_n$ are arbitrary positive integers. This means that $u_1, u_2, \ldots, u_n$, $v_1, v_2, \ldots, v_n$ in the expression (3) are formed only by those elements of V whose number of occurences in the strings of L is unlimited. Only b satisfies this requirement. It follows that in any string of L both occurrences of a and the occurence of c are terms of the string x in (3). But this implies that the intermediate terms between the occurrences of a are terms of x, hence we can find two strings y and z such that

$$x = y\, a \frown b^m \frown c\, a z.$$

The string y is obviously the null-string ω, whereas z is of the form b^p, hence

$$x = ab^m \frown c\, ab.$$

But m may be here an arbitrary positive integer. Therefore, since

$x \in L_1$, it follows that L_1 is an infinite set of strings. This fact contradicts the assumption concerning G ; L is not a contextual language and Proposition 9 is proved.

The contextual grammars may be generalized in order to generate some languages which are not context-free.

A <u>generalized contextual grammar</u> is a quadruple $G = \langle V, L_1, L_2, \mathcal{C} \rangle$, where V, L_1 and $\mathcal{C}$ have the same meaning as in the definition of a contextual grammar, whereas L_2 is a finite set of strings on the vocabulary V. We define the language L_G generated by G in the following way : L_G is a language on V and $x \in L_G$ if and only if we may express x in the form

$$x = u_1^{p_1} u_2^{p_2} \ldots u_n^{p_n} z y^p v_n^{p_n} \ldots v_2^{p_2} v_1^{p_1},$$

where $z \in L_1$, $y \in L_2$, $\langle u_i, v_i \rangle \in \mathcal{C}$ for $i = 1,2,\ldots,n$ and $p_1, p_2, \ldots, p_n$, p are positive integers such that $p_1 + p_2 \ldots + p_n = p$.

Every language generated by a generalized contextual grammar is said to be a generalized contextual language.

If, in the definition of G, we take $L_2 = \{\omega\}$, G is equivalent to a contextual grammar ; the language L_G is then precisely the language generated by the contextual grammar $\langle V, L_1, \mathcal{C} \rangle$. Indeed, the general form of a string in the contextual language generated by $\langle V, L_1, \mathcal{C} \rangle$ is

$$u_1^{p_1} u_2^{p_2} \ldots u_n^{p_n} z\, v_n^{p_n} \ldots v_2^{p_2} v_1^{p_1}$$

where $\langle u_i, v_i \rangle \in \mathcal{C}$ $(i = 1,2,\ldots,n)$ and $z \in L_1$. We have thus proved

<u>Proposition 1o</u>. <u>Every contextual language is a generalized contextual language.</u>

We may consider a contextual grammar as a particular case of generalized contextual grammar, by identifying the contextual grammar $\langle V, L_1, \mathcal{G}\rangle$ with the generalized contextual grammar $\langle V, L_1, \{\omega\}, \mathcal{G}\rangle$.

It is interesting to point out that sometimes a contextual language may be easy generated by a generalized contextual grammar which is not a contextual grammar. For instance, let us consider the language $L = \{a^n b^n\}$ $(n=1,2,\dots)$. In view of the proof of Proposition 4, L is a contextual language. We may generate L by the generalized contextual grammar (which is not a contextual grammar) $\langle V, L_1, L_2, \mathcal{G}\rangle$, where $V = \{a,b\}$, $L_1 = \{\omega\}$, $L_2 = \{b\}$, $\mathcal{G} = \{a,\omega\}$. It is known that L is not regular. We may give a similar example, with a language which is regular. In this respect let us consider the language of Curry $\{ab^n\}$. In view of Proposition 3, it is a contextual language. It is a regular language too, since it may be generated by the regular grammar containing the following two rules : $S \to Sb$ and $S \to a$. Now let us consider the generalized contextual grammar $\langle V, L_1, L_2, \mathcal{G}\rangle$, where $V = \{a,b\}$, $L_1 = \{a\}$, $L_2 = \{b\}$ $\mathcal{G} = \{\langle\omega,\omega\rangle\}$. This grammar generates the language of Curry, but it is not a contextual grammar.

Now let us show that generalized contextual languages are an effective generalization of contextual languages.

<u>Proposition 11. There exists a generalized contextual language which is not a contextual language.</u>

<u>Proof.</u> Let us consider the language $L = \{a^n b^n a^n\}$ $(n=1,2,\dots)$ It is known that this language is not context-free (see, for instance, [6], p.84). In view of Proposition 7, every contextual

language is a context-free language ; hence, L is not a contextual language. Now let us consider the generalized contextual grammar $G = \langle V, L_1, L_2, \mathcal{C} \rangle$, where $V = \{a, b\}$, $L_1 = \{\omega\}$, $L_2 = \{b\}$ and $\mathcal{C} = \{\langle a, a \rangle\}$. It is easy to see that G generates the language L.

From the proof of Proposition 11 it follows immediately:

Proposition 12. There exists a generalized contextual language which is not a context-free language.

We may now ask whether the converse of Proposition 12 is true. The answer is given by

Proposition 13. There exists a context-free language (and even a regular language) which is not a generalized contextual language.

Proof. We may consider the language $L = \{ab^m c ab^n\}$ ($m, n = 1, 2, \ldots$) used in the proof of Proposition 9. It was showed in the proof of Proposition 9 that L is regular. Let us admit that L is a generalized contextual language. Given a string x in L, its representation is of the form

$$ab^m c ab^r = u_1^{p_1} u_2^{p_2} \ldots u_n^{p_n}\, z\, y^p v_n^{p_n} \ldots v_2^{p_2} v_1^{p_1}$$

where $\langle u_i, v_i \rangle \in \mathcal{C}$ $(i = 1, \ldots, n)$, $z \in L_1$, $y \in L_2$, $p_1 + \ldots + p_n = p$ and $G = \langle V, L_1, L_2, \mathcal{C} \rangle$ is the grammar of L. By a reasoning similar to that used in the proof of Proposition 9, we find that for every positive integer m there exists a string z in L_1 such that

$$z = ab^m c ab^s,$$

where s is a non-negative integer, depending of z. But this means that L_1 contains infinitely many strings. This fact contradicts the definition of a generalized contextual grammar. It follows that L is not a generalized contextual language.

It is to be expected that every generalized contextual language is a context-sensitive language. But the construction of the corresponding context-sensitive grammar seems to be very complicated, if we think to the generation of the language $\{a^n b^n a^n\}$.

Ju.A.Šreider has introduced a new type of grammars, called neighborhood grammars (okrestnostnye grammatiki) and defined in the following way ([10]; see also [2] and [11]. Our presentation is some what different). Given a finite set V called vocabulary, two strings x and y on V, and a context $\langle u,v\rangle$ on V, we say that the pair $\{\langle u,v\rangle, y\}$ is a neighborhood of y with respect to x if we can find two strings z and w, such that $x=zuyvw$. Every pair of the form $\{\langle u,v\rangle, y\}$, where $\langle u,v\rangle$ is a context on V, whereas y is a string on V, is called a neighborhood on V. Let us consider an element θ which does not belong to V; θ will be called the boundary element. A neighborhood grammar is a triple of the form $\langle V, \theta, \mathcal{T}\rangle$, where V is a vocabulary, θ is the boundary element and $\mathcal{T}$ is a finite set of neighborhoods on the vocabulary $V \cup \{\theta\}$. Let L be a language on V. We say that L is generated by the considered neighborhood grammar if in every string x of the form $x=\theta y \theta$ (with $y \in L$)-and only in such strings - there exists in $\mathcal{T}$, for every term a_i of $x=a_1a_2\ldots a_s$, a neighborhood of a_i with respect to x.

Neighborhood grammars are closely related to the notion of context, since this notion occurs in the definition of a neighborhood. There is another notion, due to Ja.P.L.Vasilevskiĭ and

M.V.Chomjakov (see the reference in[2],p.4o), which explains this fact. Following these authors, a grammar of contexts (this name is improper, since no context occurs among its objects) is a triple $\langle V, \theta, Q \rangle$, where V and θ have the same meaning as in the definition of a neighborhood grammar, whereas Q is a finite set of strings on the vocabulary $V \cup \{\theta\}$. This grammar generates the language L on V in the following way : $x \in L$ if and only if for every string y and any strings z and w for which there exist strings u and v such that $\theta x \theta =$ $= uzywv$ we have either

1) $y = msp$, where $s \in Q$, whereas the strings m and p may be eventually null

or

2) $\theta x \theta = qrynt$, where $qr = z$, $nt = w$ and ryn is a string belonging to Q.

A string belonging to Q is said to be closed from the left (from the right) if its first (last) term is θ. A string belonging to Q is said to be closed if it is closed both from the left and from the right.

A grammar of contexts is said to be k-bounded if every non-closed string of Q is of length k, whereas every closed string of Q is of length not greater than k.

An important theorem of Borščev asserts the equivalence between languages generated by neighborhood grammars and languages generated by k-bounded grammars of contexts ([2],p.4o).

Since grammars of contexts and contextual grammars have some similarities in their definitions, it is interesting to establish more exactly the relation between them.

Proposition 14. There exists a contextual language which is regular, but which is not a neighborhood language.

Proof. Let us consider the language $L = \{a^{2n}\}$ $(n=1,2,\dots)$. This language is regular, since it is generated by the regular grammar consisting in the rules $S \longrightarrow Ta$, $T \longrightarrow Ua$, $U \longrightarrow Ta$, $T \longrightarrow a$, where S is the start symbol, $\{a\}$ is the terminal vocabulary, whereas $\{S,T,U\}$ is the non-terminal vocabulary. Let us consider the contextual grammar $G = \langle \{a\}, \{\omega\}, \{\langle a,a\rangle\}\rangle$. It is easy to see that G generates the language L; therefore L is a contextual language.

We shall show that L is not a neighborhood language. In this respect, our method will be the following. We shall consider all systems of possible neighborhoods of the terms of the string $\theta aa\theta$ and we shall show that every such system is either a system of neighborhoods of the terms of every string $\theta a^n\theta$ $(n = 2,3,4,\dots)$ or it is not a system of neighborhoods of the terms of the string $\theta a^4\theta$. It is easy to see that the first term of the string $\theta\underline{aa}\theta$ admits the following neighborhoods : 1) $\underline{\theta}$, 2) $\underline{\theta}\,\underline{a}$, 3) $\underline{\theta}\,\underline{aa}$, 4) $\underline{\theta}aa\theta$. The second term has the neighborhoods : 1) $\theta\underline{a}$, 2) $\underline{a}$, 3) $\underline{aa}$, 4) $\underline{aa}\theta$, 5) $\theta\underline{aa}$, 6) $\theta\underline{aa}\theta$. The neighborhoods of the third term are : 1) $\theta\underline{aa}$, 2) $\underline{aa}$, 3) $\underline{a}$, 4) $\underline{a}\theta$, 5) $\theta a\underline{a}\theta$, 6) $\underline{aa}\theta$. The last term has the neighborhoods : 1) $\underline{\theta}$, 2) $\underline{a\theta}$, 3) $aa\underline{\theta}$, 4) $\theta aa\underline{\theta}$. The notation $\underline{u}\underline{\underline{x}}\underline{v}$ represents hier the neighborhood $\{\langle u,v\rangle, x\}$.

It is easy to see that the fourth neighborhood of the first and of the last term cannot be a neighborhood of θ with respect to $\theta a^4\theta$. On the other hand, $\underline{a}$ is a neighborhood of a with respect to $\theta a^n\theta$ for every $n = 1,2,\dots$. It follows that no neighborhood grammar of $L = \{a^{2n}\}$ may contain one of the neighborhoods $\underline{\theta}a^2\theta$, $\theta a^2\underline{\theta}$ and $\underline{a}$. Thus, if a neighborhood grammar of L exists, it contains at least one neighborhood from every group of the following four groups of neighborhoods :

α) $\underline{\theta}$, $\underline{\theta}a$, $\underline{\theta}a^2$.

β) $\theta\underline{a}$, $\underline{a}a$, $\underline{a}a\theta$, $\theta\underline{a}a$, $\theta\underline{a}a\theta$.

γ) $\theta a\underline{a}$, $a\underline{a}$, $\underline{a}\theta$, $\theta a\underline{a}\theta$, $a\underline{a}\theta$.

δ) $\underline{\theta}$, $a\underline{\theta}$, $a^2\underline{\theta}$.

We shall consider all possible combinations between a neighborhood of the group β and a neighborhood of the group γ . By mn we shall denote the combination formed by the m-th neighborhood of β and the n-th neighborhood of γ . It is easy to see that every neighborhood grammar containing one of the combinations 12, 22, 23, 25, 42 generates a language which contains every string a^n with $n \geqslant 2$. On the other hand, every neighborhood grammar containing one of the combinations 11, 13, 14, 15, 21, 24, 31, 32, 33, 34, 35, 41, 43, 44, 45, 51, 52, 53, 54, 55 generates a language which either does not contain the string a^4 or contains every string a^n with $n \geqslant 2$. (This depends on the fact if the neighborhoods $\underline{a}a$ or $\underline{aa}$ belong or not to the considered neighborhood grammar). Thus, there exists no neighborhood grammar which generates the language $\{a^{2n}\}$.

But the definition of (generalized) contextual grammars, though adequate to the investigation of the generative power of purely contextual operations, does not correspond to the situation existing in real (natural or artificial) languages, where every string is admited only by some contexts and every context admits only some strings. Let us try to obtain a type of grammar corresponding to this more complex situation. We define a <u>contextual grammar with choice</u> as a system $G = \langle V, L_1, \mathcal{C}, \varphi \rangle$, where V, L_1 and $\mathcal{C}$ are the objects of a contextual grammar, whereas φ is a mapping defined on the universal language on V and having the values in the set of subsets of $\mathcal{C}$. We define the language generated by G as the smallest language L having the following properties : 1^o If $x \in L_1$, $x \in L$; 2^o If $y \in L$, $\langle u, v \rangle \in \varphi(y)$ and $z \in L_1$, then $uyv \in L$, $zv \in L$ and $yz \in L$. Thus, every string chooses some contexts and every context chooses some strings. We define a contextual language with choice a language which is generated by a contextual grammar with choice. The investigation of these grammars and languages would better show the generative power of contextual operations, in a manner which corresponds to the situation existing in real languages.

References

1. Y.BAR-HILLEL, M.PERLES, E.SHAMIR : On formal properties of simple phrase structure grammars. Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung, vol.14, 1961, p.143-172.
2. V.B.BORŠČEV : Okrestnostnye grammatiki. Naučno-Techničeskaja Informacija, Serija 2, 1967, No.11, p.39-41.
3. N.CHOMSKY : Three models for the description of language. IRE Transactions on Information Theory, IT-2, 3, 1956, p.113-124.
4. N.CHOMSKY : Syntactic Structures. 's Gravenhage, 1957.
5. H.CURRY : Some logical aspects of grammatical structure. Proceedings of the Symposium in Applied Mathematics, vol.12, Structure of language and its mathematical aspects, Amer. Math. Soc., 1961, p.56-68.
6. S.GINSBURG : The mathematical theory of context-free languages. McGraw-Hill Book Company, New York, 1966.
7. J.GRUŠKA : On a classification of context-free languages. Kybernetika, vol.3, no.1, 1967, p.22-29.
8. S.MARCUS : Gramatici şi automate finite. Editura Academiei R.P.R., Bucureşti, 1964.
9. S.MARCUS : Sur les grammaires à un nombre d'états fini. Cahiers de linguistique théorique et appliquée, vol.2, 1965, p.147-164
10. Ju.A.ŠREIDER : Okrestnostnaja model jazyka. Trudy Simpoziuma po primenenijam poroždajuščich grammatik. Tartu, septjabr, 1967.
11. Ju.A.ŠREIDER : Topologičeskie modeli jazyka. Vsesojuznyi Institut naučnoi i techničeskoi informacii, Moscou, 1968.