INFORMATION AND CONTROL 2, 137–167 (1959)

On Certain Formal Properties of Grammars*

NOAM CHOMSKY

Massachusetts Institute of Technology, Cambridge, Massachusetts and The Institute for Advanced Study, Princeton, New Jersey

A grammar can be regarded as a device that enumerates the sentences of a language. We study a sequence of restrictions that limit grammars first to Turing machines, then to two types of system from which a phrase structure description of the generated language can be drawn, and finally to finite state Markov sources (finite automata). These restrictions are shown to be increasingly heavy in the sense that the languages that can be generated by grammars meeting a given restriction constitute a proper subset of those that can be generated by grammars meeting the preceding restriction. Various formulations of phrase structure description are considered, and the source of their excess generative power over finite state sources is investigated in greater detail.

SECTION 1

A language is a collection of sentences of finite length all constructed from a finite alphabet (or, where our concern is limited to syntax, a finite vocabulary) of symbols. Since any language L in which we are likely to be interested is an infinite set, we can investigate the structure of L only through the study of the finite devices (grammars) which are capable of enumerating its sentences. A grammar of L can be regarded as a function whose range is exactly L. Such devices have been called "sentence-generating grammars."[1] A theory of language will contain, then, a specifica-

* This work was supported in part by the U. S. Army (Signal Corps), the U. S. Air Force (Office of Scientific Research, Air Research and Development Command), and the U. S. Navy (Office of Naval Research). This work was also supported in part by the Transformations Project on Information Retrieval of the University of Pennsylvania. I am indebted to George A. Miller for several important observations about the systems under consideration here, and to R. B. Lees for material improvements in presentation.

[1] Following a familiar technical use of the term "generate," cf. Post (1944). This locution has, however, been misleading, since it has erroneously been interpreted as indicating that such sentence-generating grammars consider language

tion of the class F of functions from which grammars for particular languages may be drawn.

The weakest condition that can significantly be placed on grammars is that F be included in the class of general, unrestricted Turing machines. The strongest, most limiting condition that has been suggested is that each grammar be a finite Markovian source (finite automaton).[2]

The latter condition is known to be too strong; if F is limited in this way it will not contain a grammar for English (Chomsky, 1956). The former condition, on the other hand, has no interest. We learn nothing about a natural language from the fact that its sentences can be effectively displayed, i.e., that they constitute a recursively enumerable set. The reason for this is clear. Along with a specification of the class F of grammars, a theory of language must also indicate how, in general, relevant structural information can be obtained for a particular sentence generated by a particular grammar. That is, the theory must specify a class Σ of "structural descriptions" and a functional Φ such that given $f \in F$ and x in the range of f, $\Phi(f,x) \in \Sigma$ is a structural description of x (with respect to the grammar f) giving certain information which will facilitate and serve as the basis for an account of how x is used and understood by speakers of the language whose grammar is f; i.e., which will indicate whether x is ambiguous, to what other sentences it is structurally similar, etc. These empirical conditions that lead us to characterize F in one way or another are of critical importance. They will not be further discussed in this paper,[3] but it is clear that we will not be able to de-

from the point of view of the speaker rather than the hearer. Actually, such grammars take a completely neutral point of view. Compare Chomsky (1957, p. 48). We can consider a grammar of L to be a function mapping the integers onto L, order of enumeration being immaterial (and easily specifiable, in many ways) to this purely syntactic study, though the question of the particular "inputs" required to produce a particular sentence may be of great interest for other investigations which can build on syntactic work of this more restricted kind.

[2] Compare Definition 9, Sec. 5.

[3] Except briefly in §*2*. In Chomsky (1956, 1957), an appropriate Φ and Σ (i.e., an appropriate method for determining structural information in a uniform manner from the grammar) are described informally for several types of grammar, including those that will be studied here. It is, incidentally, important to recognize that a grammar of a language that succeeds in enumerating the sentences will (although it is far from easy to obtain even this result) nevertheless be of quite limited interest unless the underlying principles of construction are such as to provide a useful structural description.

velop an adequate formulation of Φ and Σ if the elements of F are specified only as such "unstructured" devices as general Turing machines.

Interest in structural properties of natural language thus serves as an empirical motivation for investigation of devices with more generative power than finite automata, and more special structure than Turing machines. This paper is concerned with the effects of a sequence of increasing heavy restrictions on the class F which limit it first to Turing machines and finally to finite automata and, in the intermediate stages, to devices which have linguistic significance in that generation of a sentence automatically provides a meaningful structural description. We shall find that these restrictions are increasingly heavy in the sense that each limits more severely the set of languages that can be generated. The intermediate systems are those that assign a phrase structure description to the resulting sentence. Given such a classification of special kinds of Turing machines, the main problem of immediate relevance to the theory of language is that of determining where in the hierarchy of devices the grammars of natural languages lie. It would, for example, be extremely interesting to know whether it is in principle possible to construct a phrase structure grammar for English (even though there is good motivation of other kinds for not doing so). Before we can hope to answer this, it will be necessary to discover the structural properties that characterize the languages that can be enumerated by grammars of these various types. If the classification of generating devices is reasonable (from the point of view of the empirical motivation), such purely mathematical investigation may provide deeper insight into the formal properties that distinguish natural languages, among all sets of finite strings in a finite alphabet. Questions of this nature appear to be quite difficult in the case of the special classes of Turing machines that have the required linguistic significance.[4] This paper is devoted to a preliminary study of the properties of such special devices, viewed as grammars.

It should be mentioned that there appears to be good evidence that devices of the kinds studied here are not adequate for formulation of a full grammar for a natural language (see Chomsky, 1956, §4; 1957, Chapter 5). Left out of consideration here are what have elsewhere been

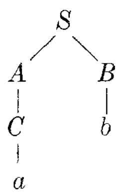
[4] In Chomsky and Miller (1958), a structural characterization theorem is stated for languages that can be enumerated by finite automata, in terms of the cyclical structure of these automata. The basic characterization theorem for finite automata is proven in Kleene (1956).

called "grammatical transformations" (Harris, 1952a, b, 1957; Chomsky, 1956, 1957). These are complex operations that convert sentences with a phrase structure description into other sentences with a phrase structure description. Nevertheless, it appears that devices of the kind studied in the following pages must function as essential components in adequate grammars for natural languages. Hence investigation of these devices is important as a preliminary to the far more difficult study of the generative power of transformational grammars (as well as, negatively, for the information it should provide about what it is in natural language that makes a transformational grammar necessary).

SECTION 2

A phrase structure grammar consists of a finite set of "rewriting rules" of the form $\varphi \rightarrow \psi$, where φ and ψ are strings of symbols. It contains a special "initial" symbol S (standing for "sentence") and a boundary symbol # indicating the beginning and end of sentences. Some of the symbols of the grammar stand for words and morphemes (grammatically significant parts of words). These constitute the "terminal vocabulary." Other symbols stand for phrases, and constitute the "nonterminal vocabulary" (S is one of these, standing for the "longest" phrase). Given such a grammar, we generate a sentence by writing down the initial string $\#S\#$, applying one of the rewriting rules to form a new string $\#\varphi_1\#$ (that is, we might have applied the rule $\#S\# \rightarrow \#\varphi_1\#$ or the rule $S \rightarrow \varphi_1$), applying another rule to form a new string $\#\varphi_2\#$, and so on, until we reach a string $\#\varphi_n\#$ which consists solely of terminal symbols and cannot be further rewritten. The sequence of strings constructed in this way will be called a "derivation" of $\#\varphi_n\#$.

Consider, for example, a grammar containing the rules: $S \rightarrow AB$, $A \rightarrow C$, $CB \rightarrow Cb$, $C \rightarrow a$, and hence providing the derivation $D = (\#S\#, \#AB\#, \#CB\#, \#Cb\#, \#ab\#)$. We can represent D diagrammatically in the form



(1)

If appropriate restrictions are placed on the form of the rules $\varphi \rightarrow \psi$ (in particular, the condition that ψ differ from φ by replacement of a single symbol of φ by a non-null string), it will always be possible to associate with a derivation a labeled tree in the same way. These trees can be taken as the structural descriptions discussed in Sec. 1, and the method of constructing them, given a derivation, will (when stated precisely) be a definition of the functional Φ. A substring x of the terminal string of a given derivation will be called a phrase of type A just in case it can be traced back to a point labeled A in the associated tree (thus, for example, the substring enclosed within the boundaries is a phrase of the type "sentence"). If in the example given above we interpret A as *Noun Phrase*, B as *Verb Phrase*, C as *Singular Noun*, a as *John*, and b as *comes*, we can regard D as a derivation of *John comes* providing the structural description (1), which indicates that *John* is a *Singular Noun* and a *Noun Phrase*, that *comes* is a *Verb Phrase*, and that *John comes* is a *Sentence*. Grammars containing rules formulated in such a way that trees can be associated with derivations will thus have a certain linguistic significance in that they provide a precise reconstruction of large parts of the traditional notion of "parsing" or, in its more modern version, immediate constituent analysis. (Cf. Chomsky (1956, 1957) for further discussion.)

The basic system of description that we shall consider is a system G of the following form: G is a semi-group under concatenation with strings in a finite set V of symbols as its elements, and I as the identity element. V is called the "vocabulary" of G. $V = V_T \cup V_N (V_T, V_N$ disjoint), where V_T is the "terminal vocabulary" and V_N the "nonterminal vocabulary." V_T contains I and a "boundary" element #. V_N contains an element S (sentence). A two-place relation $\rightarrow$ is defined on elements of G, read "can be rewritten as." This relation satisfies the following conditions:

AXIOM 1. $\rightarrow$ is irreflexive.

AXIOM 2. $A \in V_N$ if and only if there are φ, ψ, ω such that $\varphi A \psi \rightarrow \varphi \omega \psi$.

AXIOM 3. There are no φ, ψ, ω such that $\varphi \rightarrow \psi \# \omega$.

AXIOM 4. There is a finite set of pairs $(\chi_1, \omega_1), \cdots, (\chi_n, \omega_n)$ such that for all φ, ψ, $\varphi \rightarrow \psi$ if and only if there are φ_1, φ_2, and $j \leqq n$ such that $\varphi = \varphi_1 \chi_j \varphi_2$ and $\psi = \varphi_1 \omega_j \varphi_2$.

Thus the pairs (χ_j, ω_j) whose existence is guaranteed by Axiom 4 give a finite specification of the relation $\rightarrow$. In other words, we may think of the grammar as containing a finite number of rules $\chi_j \rightarrow \omega_j$ which completely determine all possible derivations.

The presentation will be greatly facilitated by the adoption of the following notational convention (which was in fact followed above).

CONVENTION 1: We shall use capital letters for strings in V_N ; small Latin letters for strings in V_T ; Greek letters for arbitrary strings; early letters of all alphabets for single symbols (members of V); late letters of all alphabets for arbitrary strings.

DEFINITION 1. $(\varphi_1, \cdots, \varphi_n)(n \geqq 1)$ is a *ψ-derivation of* ω if $\psi = \varphi_1$, $\omega = \varphi_n$, and $\varphi_i \rightarrow \varphi_{i+1} (1 \leqq i < n)$.

DEFINITION 2. A φ-derivation is *terminated* if it is not a proper initial subsequence of any φ-derivation.[5]

DEFINITION 3. The *terminal language* L_G *generated* by G is the set of strings x such that there is a terminated $\#S\#$-derivation of x.[6]

DEFINITION 4. G is *equivalent* to G^* if $L_G = L_{G^*}$.

DEFINITION 5. $\varphi \Rightarrow \psi$ if there is a φ-derivation of ψ.

$\Rightarrow$ (which is the ordinary ancestral of $\rightarrow$) is thus a partial ordering of strings in G. These notions appear, in slightly different form, in Chomsky (1956, 1957).

This paper will be devoted to a study of the effect of imposing the following additional restrictions on grammars of the type described above.

RESTRICTION 1. If $\varphi \rightarrow \psi$, then there are $A, \varphi_1, \varphi_2, \omega$ such that $\varphi = \varphi_1 A \varphi_2$, $\psi = \varphi_1 \omega \varphi_2$, and $\omega \neq I$.

RESTRICTION 2. If $\varphi \rightarrow \psi$, then there are $A, \varphi_1, \varphi_2, \omega$ such that $\varphi = \varphi_1 A \varphi_2$, $\psi = \varphi_1 \omega \varphi_2$, $\omega \neq I$, but $A \rightarrow \omega$.

RESTRICTION 3. If $\varphi \rightarrow \psi$, then there are $A, \varphi_1, \varphi_2, \omega, a, B$ such that $\varphi = \varphi_1 A \varphi_2$, $\psi = \varphi_1 \omega \varphi_2$, $\omega \neq I$, $A \rightarrow \omega$, but $\omega = aB$ or $\omega = a$.

The nature of these restrictions is clarified by comparison with Axiom 4, above. Restriction 1 requires that the rules of the grammar [i.e., the minimal pairs (χ_j, ω_j) of Axiom 4] all of be the form $\varphi_1 A \varphi_2 \rightarrow \varphi_1 \omega \varphi_2$, where A is a single symbol and $\omega \neq I$. Such a rule asserts that $A \rightarrow \omega$ in the context φ_1—φ_2 (which may be null). Restriction 2 requires that the limiting context indeed be null; that is, that the rules all be of the form $A \rightarrow \omega$, where A is a single symbol, and that each such rule may be applied independently of the context in which A appears. Restriction 3

[5] Note that a terminated derivation need not terminate in a string of V_T (i.e., it may be "blocked" at a nonterminal string), and that a derivation ending with a string of V_T need not be terminated (if, e.g., the grammar contains such rules as $ab \rightarrow cd$).

[6] Thus the terminal language L_G consists only of those strings of V_T which are derivable from $\#S\#$ but which cannot head a derivation (of $\geqq 2$ lines).

limits the rules to the form $A \rightarrow aB$ or $A \rightarrow a$ (where A,B are single nonterminal symbols, and a is a single terminal symbol).

DEFINITION 6. For $i = 1, 2, 3$, a *type i grammar* is one meeting restriction i, and a *type i language* is one with a type i grammar. A *type* 0 *grammar* (*language*) is one that is unrestricted.

Type 0 grammars are essentially Turing machines; type 3 grammars, finite automata. Type 1 and 2 grammars can be interpreted as systems of phrase structure description.

SECTION 3

Theorem 1 follows immediately from the definitions.

THEOREM 1. For both grammars and languages, type 0 $\supseteq$ type 1 $\supseteq$ type 2 $\supseteq$ type 3.

The following is, furthermore, well known.

THEOREM 2. Every recursively enumerable set of strings is a type 0 language (and conversely).[7]

That is, a grammar of type 0 is a device with the generative power of a Turing machine. The theory of type 0 grammars and type 0 languages is thus part of a rapidly developing branch of mathematics (recursive function theory). Conceptually, at least, the theory of grammar can be viewed as a study of special classes of recursive functions.

THEOREM 3. Each type 1 language is a decidable set of strings.[7a]

That is, given a type 1 grammar G, there is an effective procedure for determining whether an arbitrary string x is in the language enumerated by G. This follows from the fact that if φ_i, φ_{i+1} are successive lines of a derivation produced by a type 1 grammar, then φ_{i+1} cannot contain fewer symbols than φ_i, since φ_{i+1} is formed from φ_i by replacing a single symbol A of φ_i by a non-null string ω. Clearly any string x which has a

[7] See, for example, Davis (1958, Chap. 6, §2). It is easily shown that the further structure in type 0 grammars over the combinatorial systems there described does not affect this result.

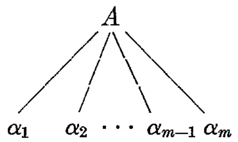
[7a] But not conversely. For suppose we give an effective enumeration of type 1 grammars, thus enumerating type 1 languages as $L_1, L_2, \cdots$. Let $s_1, s_2, \cdots$ be an effective enumeration of all finite strings in what we can assume (without restriction) to be the common, finite alphabet of $L_1, L_2, \cdots$. Given the index of a language in the enumeration $L_1, L_2, \cdots$, we have immediately a decision procedure for this language. Let M be the "diagonal" language containing just those strings s_i such that $s_i \notin L_i$. Then M is a decidable language not in the enumeration.

I am indebted to Hilary Putnam for this observation.

$\#S\#$-derivation, has a $\#S\#$-derivation in which no line repeats, since lines between repetitions can be deleted. Consequently, given a grammar G of type 1 and a string x, only a finite number of derivations (those with no repetitions and no lines longer than x) need be investigated to determine whether $x \in L_G$.

We see, therefore, that Restriction 1 provides an essentially more limited type of grammar than type 0.

The basic relation $\rightarrow$ of a type 1 grammar is specified completely by a finite set of pairs of the form $(\psi_1 A \psi_2, \psi_1 \omega \psi_2)$. Suppose that $\omega = \alpha_1 \cdots \alpha_m$. We can then associate with this pair the element



(2)

Corresponding to any derivation D we can construct a tree formed from the elements (2) associated with the transitions between successive lines of D, adding elements to the tree from the appropriate node as the derivation progresses.[8] We can thus associate a labeled tree with each derivation as a structural description of the generated sentence. The restriction on the rules $\varphi \rightarrow \psi$ which leads to type 1 grammars thus has a certain linguistic significance since, as pointed out in Sec. 1, these grammars provide a precise reconstruction of much of what is traditionally called "parsing" or "immediate constituent analysis." Type 1 grammars are the phrase structure grammars considered in Chomsky (1957, Chap. 4).

SECTION 4

Lemma 1. Suppose that G is a type 1 grammar, and X,B are particular strings of G. Let G' be the grammar formed by adding $XB \rightarrow BX$ to G. Then there is a type 1 grammar G^* equivalent to G'.

Proof. Suppose that $X = A_1 \cdots A_n$. Choose $C_1, \cdots, C_{n+1}$ new and distinct. Let Q be the sequence of rules

[8] This associated tree might not be unique, if, for example, there were a derivation containing the successive lines $\varphi_1 AB\varphi_2$, $\varphi_1 A\psi B\varphi_2$, since this step in the derivation might have used either of the rules $A \rightarrow A\psi$ or $B \rightarrow \psi B$. It is possible to add conditions on G that guarantee uniqueness without affecting the set of generated languages.

$$\begin{aligned}
A_1 \cdots A_nB &\to C_1A_2 \cdots A_nB \\
&\quad\vdots \\
&\to C_1 \cdots C_nB \\
&\to C_1 \cdots C_nC_{n+1} \\
&\to BC_2 \cdots C_{n+1} \\
&\to BA_1C_3 \cdots C_{n+1} \\
&\quad\vdots \\
&\to BA_1 \cdots A_n
\end{aligned}$$

where the left-hand side of each rule is the right-hand side of the immediately preceding rule. Let G^* be formed by adding the rules of Q to G. It is obvious that if there is a $\#S\#$-derivation of x in G^* using rules of Q, then there is a $\#S\#$-derivation of x in G^* in which the rules are applied only in the sequence Q, with no other rules interspersed (note that x is a terminal string). Consequently the only effect of adding the rules of Q to G is to permit a string $\varphi XB\psi$ to be rewritten $\varphi BX\psi$, and L_{G^*} contains only sentences of $L_{G'}$. It is clear that L_{G^*} contains all the sentences of $L_{G'}$ and that G^* meets Restriction 1.

By a similar argument it can easily be shown that type 1 languages are those whose grammars meet the condition that if $\varphi \to \psi$, then ψ is at least as long as φ. That is, weakening Restriction 1 to this extent will not increase the class of generated languages.

Lemma 2. Let L be the language containing all and only the sentences of the form $\#a^nb^ma^nb^mccc\#(m,n \geqq 1)$. Then L is a type 1 language.

Proof. Consider the grammar G with $V_T = \{a,b,c,I,\#\}$,

$$V_N = \{S, S_1, S_2, A, \bar{A}, B, \bar{B}, C, D, E, F\},$$

and the following rules:

(I) (a) $S \to CDS_1S_2F$

(b) $S_2 \to S_2S_2$

(c) $\left\{\begin{matrix} S_2F \to BF \\ S_2B \to BB \end{matrix}\right\}$

(d) $S_1 \to S_1S_1$

(e) $\left\{\begin{matrix} S_1B \to AB \\ S_1A \to AA \end{matrix}\right\}$

$$\text{(II)}\quad \text{(a)} \begin{Bmatrix} CDA \to CE\bar{A}A \\ CDB \to CE\bar{B}B \end{Bmatrix}$$

$$\text{(b)} \begin{Bmatrix} CE\bar{A} \to \bar{A}CE \\ CE\bar{B} \to \bar{B}CE \end{Bmatrix}$$

$$\text{(c)}\ E\alpha\beta \to \beta E\alpha$$

$$\text{(d)}\ E\alpha\# \to D\alpha\#$$

$$\text{(e)}\ \alpha D \to D\alpha$$

$$\text{(III)}\quad CDF\alpha \to \alpha CDF$$

$$\text{(IV)}\quad \text{(a)} \begin{Bmatrix} A, \bar{A} \to a \\ B, \bar{B} \to b \end{Bmatrix}$$

$$\text{(b)} \begin{Bmatrix} CDF\# \to CDc\# \\ CDc \to Ccc \\ Cc \to cc \end{Bmatrix}$$

where α, β range over $\{A, B, F\}$.

It can now be determined that the only $\#S\#$-derivations of G that terminate in strings of V_T are produced in the following manner:

(1) the rules of (I) are applied as follows: (a) once, (b) $m - 1$ times for some $m \geqq 1$, (c) m times, (d) $n - 1$ times for some $n \geqq 1$, and (e) n times, giving

$$\#CD\alpha_1 \cdots \alpha_{n+m}F\#$$

where $\alpha_i = A$ for $i \leqq n$, $\alpha_i = B$ for $i > n$

(2) the rules of (II) are applied as follows: (a) once and (b) once, giving

$$\#\bar{\alpha}_1 CE\alpha_1 \cdots \alpha_{n+m}F\#$$ [9]

(c) $n + m$ times and (d) once, giving

$$\#\bar{\alpha}_1 C\alpha_2 \cdots \alpha_{n+m}FD\alpha_1\#$$

(e) $n + m$ times, giving

$$\#\bar{\alpha}_1 CD\alpha_2 \cdots \alpha_n F\alpha_1\#$$

(3) the rules of (II) are applied, as in (2), $n + m - 1$ more times, giving

$$\#\bar{\alpha}_1 \cdots \bar{\alpha}_{n+m}CDF\alpha_1 \cdots \alpha_{n+m}\#$$

[9] Where here and henceforth, $\bar{\alpha}_i = \bar{A}$ if $\alpha_i = A$, $\bar{\alpha}_i = \bar{B}$ if $\alpha_i = B$. Note that use of rules of the type of (II), (b), (c), (e), and (III) is justified by Lemma 1.

(4) the rule (III) is applied $n + m$ times, giving

$$\#\bar{\alpha}_1 \cdots \bar{\alpha}_{n+m}\alpha_1 \cdots \alpha_{n+m}CDF\#$$

(5) the rules of (IV) are applied, (a) $2(n + m)$ times, (b) once, giving

$$\#a^n b^m a^n b^m ccc\#$$

Any other sequence of rules (except for a certain freedom in point of application of [IVa]) will fail to produce a derivation terminating in a string of V_T. Notice that the form of the terminal string is completely determined by step (1) above, where n and m are selected. Rules (II) and (III) are nothing but a copying device that carries any string of the from $\#CDXF\#$ (where X is any string of A's and B's) to the corresponding string $\#XXCDF\#$, which is converted by (IV) into terminal form.

By Lemma 1, there is a type 1 grammar G^* equivalent to G, as was to be proven.

THEOREM 4. There are type 1 languages which are not type 2 languages.

PROOF. We have seen that the language L consisting of all and only the strings $\#a^n b^m a^n b^m ccc\#$ is a type 1 language. Suppose that G is a type 2 grammar of L. We can assume for each A in the vocabulary of G that there are infinitely many x's such that $A \Rightarrow x$ (otherwise A can be eliminated from G in favor of a finite number of rules of the form $B \rightarrow \varphi_1 z \varphi_2$ whenever G contains the rule $B \rightarrow \varphi_1 A \varphi_2$ and $A \Rightarrow z$). L contains infinitely many sentences, but G contains only finitely many symbols. Therefore we can find an A such that for infinitely many sentences of L there is an $\#S\#$-derivation the next-to-last line of which is of the form xAy (i.e., A is its only nonterminal symbol). From among these, select a sentence $s = \#a^n b^m a^n b^m ccc\#$ such that $m + n > r$, where $a_1 \cdots a_r$ is the longest string z such that $A \rightarrow z$ (note that there must be a z such that $A \rightarrow z$, since A appears in the next-to-last line of a derivation of a terminal string; and, by Axiom 4, there are only finitely many such z's). But now it is immediately clear that if $(\varphi_1, \cdots, \varphi_{t+1})$ is a $\#S\#$-derivation of s for which $\varphi_t = \#xAy\#$, then no matter what x and y may be,

$$(\varphi_1, \cdots, \varphi_t)$$

is the initial part of infinitely many derivations of terminal strings not in L. Hence G is not a grammar of L.

We see, therefore, that grammars meeting Restriction 2 are essentially less powerful than those meeting only Restriction 1. However, the extra power of grammars that do not meet Restriction 2 appears, from the above results, to be a defect of such grammars, with regard to the intended interpretation. The extra power of type 1 grammars comes (in part, at least) from the fact that even though only a single symbol is rewritten with each addition of a new line to a derivation, it is nevertheless possible in effect to incorporate a permutation such as $AB \rightarrow BA$ (Lemma 1). The purpose of permitting only a single symbol to be rewritten was to permit the construction of a tree (as in Sec. 2) as a structural description which specifies that a certain segment x of the generated sentence is an A (e.g., in the example in Sec. 2, *John* is a *Noun Phrase*). The tree associated with a derivation such as that in the proof of Lemma 1 will, where it incorporates a permutation $AB \rightarrow BA$, specify that the segment derived ultimately from the B of $\cdots BA \cdots$ is an A, and the segment derived from the A of $\cdots BA \cdots$ is a B. For example, a type 1 grammar in which both *John will come* and *will John come* are derived from an earlier line *Noun Phrase–Modal–Verb*, where *will John come* is produced by a permutation, would specify that *will* in *will John come* is a *Noun Phrase* and *John* a *Modal*, contrary to intention. Thus the extra power of type 1 grammars is as much a defect as was the still greater power of unrestricted Turing machines (type 0 grammars).

A type 1 grammar may contain minimal rules of the form $\varphi_1 A \varphi_2 \rightarrow \varphi_1 \omega \varphi_2$, whereas in a type 2 grammar, φ_1 and φ_2 must be null in this case. A rule of the type 1 form asserts, in effect, that $A \rightarrow \omega$ in the context $\varphi_1 — \varphi_2$. Contextual restrictions of this type are often found necessary in construction of phrase structure descriptions for natural languages. Consequently the extra flexibility permitted in type 1 grammars is important. It seems clear, then, that neither Restriction 1 nor Restriction 2 is exactly what is required for the complete reconstruction of immediate constituent analysis. It is not obvious what further qualification would be appropriate.

In type 2 grammars, the anomalies mentioned in footnote 5 are avoided. The final line of each terminated derivation is a string in V_T, and no string in V_T can head a derivation of more than one line.

SECTION 5

We consider now grammars meeting Restriction 2.

DEFINITION 7. A grammar is *self-embedding* (s.e.) if it contains an A such that for some $\varphi, \psi (\varphi \neq I \neq \psi)$, $A \Rightarrow \varphi A \psi$.

DEFINITION 8. A grammar G is *regular* if it contains only rules of the form $A \rightarrow a$ or $A \rightarrow BC$, where $B \neq C$; and if whenever $A \rightarrow \varphi_1 B \varphi_2$ and $A \rightarrow \psi_1 B \psi_2$ are rules of G, then $\varphi_i = \psi_i (i = 1, 2)$.

THEOREM 5. If G is a type 2 grammar, there is a regular grammar G^* which is equivalent to G and which, furthermore, is non-s.e. if G is non-s.e.

PROOF. Define $L(\varphi)$ (i.e., length of φ) to be m if $\varphi = \alpha_1 \cdots \alpha_m$, where $\alpha_i \neq I$.

Given a type 2 grammar G, consider all derivations $D = (\varphi_1, \cdots, \varphi_t)$ meeting the following four conditions:

(a) for some A, $\varphi_1 = A$

(b) D contains no repeating lines

(c) $L(\varphi_{t-1}) < 4$

(d) $L(\varphi_t) \geqq 4$ or φ_t is terminal.

Clearly there is a finite number of such derivations. Let G_1 be the grammar containing the minimal rule $\varphi \rightarrow \psi$ just in case for some such derivation D, $\varphi = \varphi_1$ and $\psi = \varphi_t$. Clearly G_1 is a type 2 grammar equivalent to G, and is non-s.e. if G is non-s.e., since $\varphi \rightarrow \psi$ in G_1 only if $\varphi \Rightarrow \psi$ in G.

Suppose that G_1 contains rules R_1 and R_2:

$$R_1 : A \rightarrow \varphi_1 B \varphi_2 = \omega_1 \omega_2 \omega_3 \omega_4 (\omega_i \neq I)$$

$$R_2 : A \rightarrow \psi_1 B \psi_2$$

where $\varphi_1 \neq \psi_1$ or $\varphi_2 \neq \psi_2$. Replace R_1 by the three rules

$$R_{11} : A \rightarrow CD$$

$$R_{12} : C \rightarrow \omega_1 \omega_2$$

$$R_{13} : D \rightarrow \omega_3 \omega_4$$

where C and D are new and distinct. Continuing in this way, always adding new symbols, form G_2 equivalent to G_1, non-s.e. if G_1 is non-s.e., and meeting the second of the regularity conditions.

If G_2 contains a rule $A \rightarrow \alpha_1 \cdots \alpha_n (\alpha_i \neq I, n > 2)$, replace it by the rules

$$R_1 : A \rightarrow \alpha_1 \cdots \alpha_{n-2} B$$

$$R_2 : B \rightarrow \alpha_{n-1} \alpha_n$$

where B is new. Continuing in this way, form G_3.

If G_3 contains $A \rightarrow ab (a \neq I \neq b)$, replace it by $A \rightarrow BC$, $B \rightarrow a$, $C \rightarrow b$, where B and C are new. If G_3 contains $A \rightarrow aB$, replace it by

$A \to CB$, $C \to a$, where C is new. If it contains $A \to Ba$, replace this by $A \to BC$, $C \to a$, where C is new. Continuing in this way form G_4. G_4 then is the grammar G^* required for the theorem.

Theorem 5 asserts in particular that all type 2 languages can be generated by grammars which yield only trees with no more than two branches from each node. That is, from the point of view of generative power, we do not restrict grammars by requiring that each phrase have at most two immediate constituents (note that in a regular grammar, a "phrase" has one immediate constituent just in case it is interpreted as a word or morpheme class, i.e., a lowest level phrase; an immediate constituent in this case is a member of the class).

DEFINITION 9. Suppose that Σ is a finite state Markov source with a symbol emitted at each inter-state transition; with a designated initial state S_0 and a designated final state S_f; with # emitted on transition from S_0 and from S_f to S_0, and nowhere else; and with no transition from S_f except to S_0. Define a sentence as a string of symbols emitted as the system moves from S_0 to a first recurrence of S_0. Then the set of sentences that can be emitted by Σ is a *finite state language.*[10]

Since Restriction 3 limits the rules to the form $A \to aB$ or $A \to a$, we immediately conclude the following.

THEOREM 6. The type 3 languages are the finite state languages.

PROOF. Suppose that G is a type 3 grammar. We interpret the symbols of V_N as designations of states and the symbols of V_T as transition symbols. Then a rule of the form $A \to aB$ is interpreted as meaning that a is emitted on transition from A to B. An $\#S\#$-derivation of G can involve only one application of a rule of the form $A \to a$. This can be interpreted as indicating transition from A to a final state with a emitted. The fact that # bounds each sentence of L_G can be understood as indicating the presence of an initial state S_0 with # emitted on transition from S_0 to S, and as a requirement that the only transition from the final state is to S_0, with # emitted. Thus G can be interpreted as a system of the type described in Definition 9. Similarly, each such system can be described as a type 3 grammar.

[10] Alternatively, Σ can be considered as a finite automaton, and the generated finite state language, as the set of input sequences that carry it from S_0 to a first recurrence of S_0. Cf. Chomsky and Miller (1958) for a discussion of properties of finite state languages and systems that generate them from a point of view related to that of this paper. A finite state language is essentially what is called in Kleene (1956) a "regular event."

Restriction 3 limits the rules to the form $A \rightarrow aB$ or $A \rightarrow a$. From Theorem 5 we see that Restriction 2 amounts to a limitation of the rules to the form $A \rightarrow aB$, $A \rightarrow a$, or $A \rightarrow BC$ (with the first type dispensable). Hence the fundamental feature distinguishing type 2 grammars (systems of phrase structure) from type 3 grammars (finite automata) is the possibility of rules of the form $A \rightarrow BC$ in the former. This leads to an important difference in generative power.

THEOREM 7. There exist type 2 languages that are not type 3 languages. (Cf. Chomsky, 1956, 1957.)

In Chomsky (1956), three examples of non-type 3 languages were presented. Let L_1 be the language containing just the strings $a^n b^n$; L_2, the language containing just the strings xy, where x is a string of a's and b's and y is the mirror image of x; L_3, the language consisting of all strings xx where x is a string of a's and b's. Then L_1, L_2, and L_3 are not type 3 languages. L_1 and L_2 are type 2 languages (cf. Chomsky, 1956). L_3 is a type 1 language but not a type 2 language, as can be shown by proofs similar to those of Lemma 2 and Theorem 4.[11]

Suppose that we extend the power of a finite automaton by equipping it with a finite number of counters, each of which can assume infinitely many positions. We permit each counter to shift position in a fixed way with each inter-state transition, and we permit the next transition to be determined by the present state and the present readings of the counters. A language generated (as in Definition 9) by a system of this sort (where each counter begins in a fixed position) will be called a *counter language*. Clearly L_1, though not a finite state (type 3) language, is a counter language. Several different systems of this general type are studied by Schützenberger, (1957), where the following, in particular, is proven.

THEOREM 8. L_2 is not a counter language.

Thus there are type 2 languages that are not counter languages.[12] To summarize, L_1 is a counter language and a type 2 language, but not a type 3 (finite state) language; L_2 is a type 2 language but not a counter language (hence not a type 3 language); and L_3 is a type 1 language but not a type 2 language.

[11] In Chomsky (1956, p. 119) and Chomsky (1957, p. 34), it was erroneously stated that L_3 cannot be generated by a phrase structure system. This is true for a type 2, but not a type 1 phrase structure system.

[12] The further question whether all counter languages are type 2 languages (i.e., whether counter languages constitute a step between types 2 and 3 in the hierarchy being considered here) has not been investigated.

From Theorems 2, 3, 4 and 7, we conclude:

THEOREM 9. Restrictions 1, 2 and 3 are increasingly heavy. That is, the inclusion in Theorem 1 is proper inclusion, both for grammars (trivially) and for languages.

The fact that L_2 is a type 2 language but neither a type 3 nor a counter language is important, since English has the essential properties of L_2 (Chomsky, 1956, 1957). We can conclude from this that finite automata (even with a finite number of infinite counters) that produce sentences from "left to right" in the manner of Definition 9 cannot constitute the class F (cf. Sec. 1) from which grammars are drawn; i.e., the devices that generate language cannot be of this character.

SECTION 6

The importance of gaining a better understanding of the difference in generative power between phrase structure grammars and finite state sources is clear from the considerations reviewed in Sec. 5. We shall now show that the source of the excess of power of type 2 grammars over type 3 grammars lies in the fact that the former may be self-embedding (Definition 7). Because of Theorem 5 we can restrict our attention to regular grammars.

Construction: Let G be a non-s.e. regular (type 2) grammar. Let

$$K = \{(A_1, \cdots, A_m) \mid m = 1 \quad \text{or,} \quad \text{for} \quad 1 \leqq i < j \leqq m, A_i \rightarrow \varphi A_{i+1} \psi \quad \text{and} \quad A_i \neq A_j\}.$$

We construct the grammar G' with each nonterminal symbol represented in the form $[B_1 \cdots B_n]_i (i = 1, 2)$, where the B_j's are in turn nonterminal symbols of G, as follows:[13]

Suppose that $(B_1, \cdots, B_n) \in K$.

(i) If $B_n \rightarrow a$ in G, then $[B_1 \cdots B_n]_1 \rightarrow a[B_1 \cdots B_n]_2$.

(ii) If $B_n \rightarrow CD$ where $C \neq B_i \neq D (i \leqq n)$, then

(a) $[B_1 \cdots B_n]_1 \rightarrow [B_1 \cdots B_n C]_1$

(b) $[B_1 \cdots B_n C]_2 \rightarrow [B_1 \cdots B_n D]_1$

(c) $[B_1 \cdots B_n D]_2 \rightarrow [B_1 \cdots B_n]_2$.

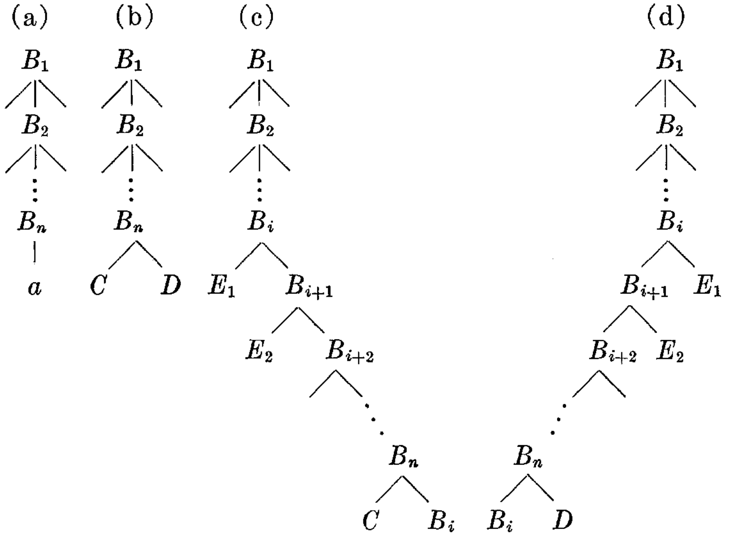
[13] Since the nonterminal symbols of G and G' are represented in different forms, we can use the symbols $\rightarrow$ and $\Rightarrow$ for both G and G' without ambiguity.

(iii) If $B_n \to CD$ where $B_i = D$ for some $i \leqq n$, then
(a) $[B_1 \cdots B_n]_1 \to [B_1 \cdots B_n C]_1$
(b) $[B_1 \cdots B_n C]_2 \to [B_1 \cdots B_i]_1$.

(iv) If $B_n \to CD$ where $B_i = C$ for some $i \leqq n$, then
(a) $[B_1 \cdots B_i]_2 \to [B_1 \cdots B_n D]_1$
(b) $[B_1 \cdots B_n D]_2 \to [B_1 \cdots B_n]_2$.

We shall prove that G' is equivalent to G (when slightly modified).

The character of this construction can be clarified by consideration of the trees generated by a grammar (cf. Sec. 3). Since G is regular and non-s.e., we have to consider only the following configurations:

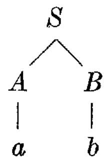


(3)

where at most two of the branches proceeding from a given node are non-null; in case (b), no node dominated by B_n is labeled $B_i (i \leqq n)$; and in each case, $B_1 = S$.

(i) of the construction corresponds to case (3a), (ii) to (3b), (iii) to (3c), and (iv) to (3d). (3c) and (3d) are the only possible kinds of recursion. If we have a configuration of the type (3c), we can have substrings of the form $(x_1 \cdots x_{n-i} y)^k$ (where $E_j \Rightarrow x_j$, $C \Rightarrow y$) in the resulting terminal strings. In the case of (3d) we can have substrings of the form $(y x_{n-i} \cdots x_1)^k$ (where $D \Rightarrow y$, $E_j \Rightarrow x_j$). (iii) and (iv) accommodate these possibilities by permitting the appropriate cycles in G'. To the earliest (highest) occurrence of a particular nonterminal symbol

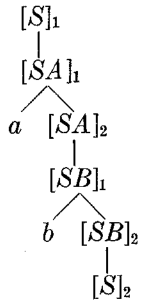
B_n in a particular branch of a tree, the construction associates two non-terminal symbols $[B_1 \cdots B_n]_1$ and $[B_1 \cdots B_n]_2$, where $B_1, \cdots, B_{n-1}$ are the labels of the nodes dominating this occurrence of B_n. The derivation in G' corresponding to the given tree will contain a subsequence $(z[B_1 \cdots B_n]_1, \cdots zx[B_1 \cdots B_n]_2)$, where $B_n \Rightarrow x$ and z is the string preceding this occurrence of x in the given derivation in G. For example, corresponding to a tree of the form



(4)

generated by a grammar G, the corresponding G' will generate the derivation (5) with the accompanying tree:

1.	$[S]_1$	
2.	$[SA]_1$	(iia)
3.	$a[SA]_2$	(i)
4.	$a[SB]_1$	(iib)
5.	$ab[SB]_2$	(i)
6.	$ab[S]_2$	(iic)



(5)

where the step of the construction permitting each line is indicated at the right.

We now proceed to establish that the grammar G' given by this construction is actually equivalent (with slight modification) to the given grammar G. This result, which requires a long sequence of introductory lemmas, is stated in a following paragraph as Theorem 10. From this we will conclude that given any non-s.e. type 2 grammar, we can construct an equivalent type 3 grammar (with many vacuous transitions which are, however, eliminable; cf. Chomsky and Miller, 1958). From this follows the main result of the paper (Theorem 11), namely, that the extra power of phrase structure grammars over finite automata as language generators lies in the fact that phrase structure grammars may be self-embedding.

LEMMA 3. If $(A_1, \cdots, A_m) \in K$, where K is as in the construction, then $A_j \Rightarrow \varphi A_k \psi$, for $1 \leqq j \leqq k \leqq m$.

LEMMA 4. If $[B_1 \cdots B_n]_i \rightarrow x[B_1 \cdots B_m]_j$, $C \neq B_k (k \leqq m, n)$, and $C \rightarrow \alpha B_1 \beta$, then $[CB_1 \cdots B_n]_i \rightarrow x[CB_1 \cdots B_m]_j$.

Proofs are immediate.

LEMMA 5. If $(B_1, \cdots, B_n) \in K$ and $1 < m \leqq n$, then

(a) if $B_m \Rightarrow \varphi B_1$, it is not the case that $B_i \Rightarrow B_m \psi (i \leqq n; i \neq m)$

(b) if $B_m \Rightarrow B_1 \varphi$, it is not the case that $B_i \Rightarrow \psi B_m (i \leqq n; i \neq m)$

(c) if $B_m \Rightarrow \varphi B_1 \psi$, it is not the case that $B_i \Rightarrow \omega_1 B_m \omega_2 B_m \omega_3 (i \leqq n)$

PROOF. Suppose that $B_m \Rightarrow \varphi B_1$ and for some $i \neq m$, $B_i \Rightarrow B_m \psi$. $\therefore \varphi \neq I \neq \psi$. By lemma 3, $B_1 \Rightarrow \omega_1 B_i \omega_2$. $\therefore B_m \Rightarrow \varphi \omega_1 B_i \omega_2 \Rightarrow \varphi \omega_1 B_m \psi \omega_2$. Contra., since now B_m is self-embedded. Similarly, case (b). Suppose $B_m \Rightarrow \varphi B_1 \psi$ and for some i, $B_i \Rightarrow \omega_1 B_m \omega_2 B_m \omega_3$. $\therefore B_1 \Rightarrow \chi_1 B_i \chi_2 \Rightarrow \chi_1 \omega_1 B_m \omega_2 B_m \omega_3 \chi_2 \Rightarrow \omega_4 B_1 \omega_5 B_1 \omega_6 \Rightarrow \omega_7 B_1 \omega_8 B_1 \omega_9 B_1 \omega_6$. Contra. (s.e.).

To facilitate proofs, we adopt the following notational convention:

CONVENTION 2. Suppose that $(\varphi_1, \cdots, \varphi_r)$ is a derivation in G' formed by construction. Then $\varphi_i = a_1 \cdots a_i Q_i$ (where Q_i is the unique non-terminal symbol that can appear in a derivation[14]), $Q_i \rightarrow a_{i+1} Q_{i+1}$.[15] $z_n^m = a_m \cdots a_n$. $z_n = z_n^1$.

LEMMA 6. Suppose that $D = (\varphi_1, \cdots, \varphi_r)$ is a derivation in G' where $Q_r = [B_1]_2$. Then:

(I) if $\varphi_1 = [B_1]_1$, $(C_1, \cdots, C_{m+1}) \in K$, $C_i \rightarrow A_{i+1} C_{i+1}$ (for $1 \leqq i \leqq m$), and $C_{m+1} = B_1$, then there is a derivation

$$([C_1 \cdots C_m B_1]_1, \cdots, z_r[C_1]_2) \quad \text{in } G'.$$

(II) if $\varphi_1 = [B_1 \cdots B_n]_1$ and $B_n \Rightarrow x B_1$, then there is a derivation

$$([B_n]_1, \cdots, z_r[B_n]_2) \quad \text{in } G'.$$

PROOF. Proof is by simultaneous induction on the length of z_r, i.e., the number of non-null symbols among $a_1, \cdots, a_r$.

Suppose that the length of z_r is 1. $\therefore$ there is one and only one i s.t. $Q_i = [\cdots]_1$ and $Q_{i+1} = [\cdots]_2$.

(a) Suppose that $i > 1$. Then $\varphi_i = Q_i$ is formed from Q_{i-1} by a rule whose source is (iia) or (iiia), and $\varphi_{i+2} = a_{i+1} Q_{i+2}$ is formed from $\varphi_{i+1} = a_{i+1} Q_{i+1}$ by a rule whose source is (iic) or (ivb). But for some

[14] Unless the initial line contained more than one nonterminal symbol, a case which will never arise below.

[15] Note that a_{i+1} will always be I unless the step of the construction justifying $\varphi_i \rightarrow \varphi_{i+1}$ is (i). a_1 will generally be I in this sequence of theorems.

k, $Q_{i-1} = [B_1 \cdots B_k]_1$, $Q_i = [B_1 \cdots B_{k+1}]_1$, $Q_{i+1} = [B_1 \cdots B_{k+1}]_2$, $Q_{i+2} = [B_1 \cdots B_k]_2$. $\therefore$ $B_k \to B_{k+1}D$ for some D, $B_k \to EB_{k+1}$, for some E, which contradicts the assumption that G is regular. $\therefore$ $i = 1$.

(b) Consider now (I). Since $i = 1$, $r = 2$. $\therefore B_1 \to z_2$. By assumption about the C_i's and m applications of Lemma 4, and (i) of the construction, $[C_1 \cdots C_mB_1]_1 \to z_2[C_1 \cdots C_mB_1]_2$. Since $C_i \to A_{i+1}C_{i+1}(C_i \neq C_j$ for $1 \leqq i < j \leqq m + 1$, since $(C_1, \cdots, C_{m+1}) \in K$ by assumption), it follows that $[C_1 \cdots C_mB_1]_2 \to [C_1 \cdots C_m]_2 \to [C_1 \cdots C_{m-1}]_2 \cdots \to [C_1]_2$. $\therefore$ $([C_1 \cdots C_mB_1]_1, z_2[C_1 \cdots C_mB_1]_2, z_2[C_1 \cdots C_m]_2, \cdots, z_2[C_1]_2)$ is the required derivation.

(c) Consider now (II). Since $i = 1$, $B_n \to z_2$ and $[B_n]_1 \to z_2[B_n]_2$, by (i) of construction. $\therefore ([B_n]_1, z_2[B_n]_2)$ is the required derivation.

This proves the lemma for the case of z_r of length 1.

Suppose it is true in all cases where z_r is of length $< t$.

Consider (I). Let D be such that z_r is of length t. If none of $C_1, \cdots, C_m$ appears in any of the Q_i's in D, then the proof is just like (a), above. Suppose that φ_j is the earliest line in which one of $C_1, \cdots, C_m$, say C_k, appears in Q_j. $j > 1$, since $C_1, \cdots, C_m \neq B_1$. By assumption of non-s.e., the rule $Q_{j-1} \to a_jQ_j$ used to form φ_j can only have been introduced by (iib).[16] $\therefore$ $Q_{j-1} = [B_1 \cdots B_nE]_2$, $Q_j = [B_1 \cdots B_nC_k]_1$, $B_n \to EC_k$.

But $C_1, \cdots, C_m$ do not occur in $Q_1, \cdots, Q_{j-1}$ and

$$(C_1, \cdots, C_m, B_1) \in K.$$

$\therefore$, by Lemma 4,

$$([C_1 \cdots C_mB_1]_1, \cdots, z_{j-1}[C_1 \cdots C_mB_1 \cdots B_nE]_2) \tag{6}$$

is a derivation. Furthermore z_{j-1} is not null, since there is at least one transition from $[\cdots]_1$ to $[\cdots]_2$ in (6), which must therefore have been introduced by (i) of the construction. But $B_n \to EC_k$. $\therefore$

$$[C_1 \cdots C_mB_1 \cdots B_nE]_2 \to [C_1 \cdots C_k]_1 \tag{7}$$

[by (iiib)]. Furthermore we know that

$$([B_1 \cdots B_nC_k]_1, \cdots, z_r^{j+1}[B_1]_2) \tag{8}$$

[16] It can only have been introduced by (iia), (iib), (iiia), (iva), or C_k will appear in Q_{j-1}. Suppose (iia). $\therefore$ $Q_{j-1} = [B_1 \cdots B_q]_1$, $Q_j = [B_1 \cdots B_qC_k]_1$, and $B_q \to C_kD$. But $C_k \Rightarrow \psi B_1$. Contra. by Lemma 5 (a). Suppose (iiia). Same. Suppose (iva). $\therefore$ $Q_{j-1} = [B_1 \cdots B_i]_2$, $Q_j = [B_1 \cdots B_{i+q}]_1$ $(q \geqq 1)$, $B_{i+q-1} \to B_iB_{i+q}$, where $C_k = B_{i+s}$ $(1 \leqq s \leqq q)$. But $C_k \Rightarrow \psi B_1$, $\psi \neq I$. $\therefore C_k \Rightarrow \psi\omega_1B_{i+q-1}\omega_2 \to \psi\omega_1B_iB_{i+q}\omega_2 \Rightarrow \psi\omega_3C_k\omega_4B_{i+q}\omega_2$, contra. $\therefore$ introduced by (iib).

must be a derivation, since $[B_1 \cdots B_n C_k]_1 = Q_j$; i.e., (8) is just the tail end $(\varphi_j, \cdots, \varphi_r)$ of D, with the initial segment z_j deleted from each of $\varphi_j, \cdots, \varphi_r$. Since z_{j-1} is not null, z_r^{j+1} is shorter than z_r, hence is of length $<t$. Also, $C_k \Rightarrow xB_1$, by assumption. $\therefore$ by inductive hypothesis (II), there is a derivation

$$([C_k]_1, \cdots, z_r^{j+1}[C_k]_2) \tag{9}$$

$\therefore$ by inductive hypothesis (I), there is a derivation

$$([C_1 \cdots C_k]_1, \cdots, z_r^{j+1}[C_1]_2) \tag{10}$$

Combining (6), (7), (10), we have the required derivation.

Consider now (II). If $n = 1$ or there is no such derivation of length t, the proof is trivial. Assume $n > 1$.

Let φ_j contain the first Q of the form $[B_1 \cdots B_m]_1 (j > 1, m \leqq n)$. Since $B_n \Rightarrow xB_1$, it follows from Lemmas 3, 5 that $B_m \Rightarrow yB_1$. Since $m \leqq n$, we see by checking through the possibilities in the construction that not all of $Q_1, \cdots, Q_{j-1}$ are of the form $[\cdots]_1$. $\therefore$ there was at least one application of (i) in forming $(\varphi_1, \cdots, \varphi_{j-1})$. $\therefore$ z_{j-1} is not null. But

$$([B_1 \cdots B_m]_1, \cdots, z_r^{j+1}[B_1]_2) \tag{11}$$

is, like (8), a derivation. $\therefore$ by inductive hypothesis (II), there is a derivation

$$([B_m]_1, \cdots, z_r^{j+1}[B_m]_2) \tag{12}$$

where z_r^{j+1} is shorter than z_r.

Let φ_k contain the first Q of the form $[B_1 \cdots B_m]_2 (m \leqq n)$. As above, $B_m \Rightarrow yB_1$. From Lemma 5 it follows that the rule used to form φ_{k+1} must be justified by (iic) or (ivb) of the construction. In either case, $Q_{k+1} = [B_1 \cdots B_{m-1}]_2$. Similarly, we show that

$$([B_1 \cdots B_m]_2, \cdots, [B_1]_2) \tag{13}$$

is a derivation. $\therefore$ $z_r = z_k$.

Let $q = \min(j,k)$. Then all of $Q_2, \cdots, Q_{q-1}$ are of the form

$$[B_1 \cdots B_{n+s}]_i .$$

It is clear that we can construct $\psi_1, \cdots, \psi_{q-1}$ s.t. for $p < q$, $\psi_p = z_p Q_p'$, where $Q_p' = [B_n \cdots B_{n+v}]_i$ when $Q_p = [B_1 \cdots B_{n+v}]_i$. Consequently

$$([B_n]_1, \cdots, z_{q-1}Q'_{q-1}) \tag{14}$$

is a derivation.

Suppose $q = j$. $\therefore$ $Q_{q-1} = [B_1 \cdots B_{n+v}]_i \rightarrow a_j[B_1 \cdots B_m]_1 = Q_j$, where $m \leqq n < n + v$. $\therefore$ this rule can only have been introduced by (iiib) of the construction. $\therefore$ $i = 2$ and $B_{n+v-1} \rightarrow B_{n+v}B_m$.

Case 1. Suppose $m = n$. $\therefore$

$$[B_n \cdots B_{n+v}]_2 \rightarrow [B_n]_1 = [B_m]_1 \tag{15}$$

Combining (14), (15), (12), We have the required derivation.

Case 2. Suppose $m < n$. $\therefore$ $B_m \neq B_n, \cdots, B_{n+v}$. $\therefore$

$$[B_n \cdots B_{n+v}]_2 \rightarrow [B_n \cdots B_{n+v-1}B_m]_1 \tag{16}$$

by (iib). We have seen that $B_m \Rightarrow yB_1$. $\therefore$ $B_{n+v-1}\ w \Rightarrow B_1$. $\therefore$ for $s < v - 1$, $B_{n+s} \rightarrow E_sB_{n+s+1}$, by Lemma 5.

But (12) is a derivation where z_r^{j+1} is of length $<t$. $\therefore$ by inductive hypothesis (I) there is a derivation

$$([B_n \cdots B_{n+v-1}B_m]_1, \cdots, z_r^{j+1}[B_n]_2) \tag{17}$$

Combining (14), (16), (17), we have the required derivation.

Suppose, finally, that $q = k$. We have seen that in this case $z_r = z_k$. But $Q_{q-1} = [B_1 \cdots B_{n+v}]_i \rightarrow a_k[B_1 \cdots B_m]_2$, where $m \leqq n < n + v$. $\therefore$ this rule can only have been introduced by (iic) or (ivb). In either case, $i = 2$, $m = n$, $v = 1$, and $Q'_{q-1} = [B_nB_{n+v}]_2 \rightarrow a_k[B_n]_2$. Combining this with (14) we have the required derivation.

We have thus shown that the lemma is true in case z_r is of length 1, and that it is true for z_r of length t on the assumption that it holds for z_r of length $<t$. Therefore it holds for every derivation D.

LEMMA 7. Suppose that $D = (\varphi_1, \cdots, \varphi_r)$ is a derivation in G' where $Q_1 = [B_1]_1$. Then

(I) if $\varphi_r = [B_1]_2$, $(C_1, \cdots, C_{m+1}) \in K$, $C_i \rightarrow C_{i+1}A_{i+1}$ $(1 \leqq i \leqq m)$, and $C_{m+1} = B_1$, then there is a derivation

$$([C_1]_1, \cdots, z_r[C_1 \cdots C_mB_1]_2) \quad \text{in } G'.$$

(II) if $\varphi_r = [B_1 \cdots B_n]_2$ and $B_n \Rightarrow B_1x$, then there is a derivation

$$([B_n]_1, \cdots, z_r[B_n]_2) \quad \text{in } G'.$$

The proof is analogous to that of Lemma 6. In the inductive step, case (I), we take Q_j as the *last* of the Q's in which one of $C_1, \cdots, C_m$

appears, and instead of (iiib) in (7), we form

$$[C_1 \cdots C_k]_2 \to [C_1 \cdots C_m B_1 \cdots B_n E]_1$$

by (iva). The proof goes through as above, with similar modifications throughout. In case (II) of the inductive step we let Q_j be the *last* Q of the form $[B_1 \cdots B_m]_2 (j < r, m \leqq n)$, and Q_k the *last* Q of the form $[B_1 \cdots B_m]_1 (m \leqq n)$. Taking $q = \max(j,k)$ [instead of $\min(j,k)$], the proof is analogous throughout, with (iva) taking the place of (iiib).

In general, because of the symmetries in case (iii), (iv) of the construction [reflecting the parallel possibilities (3c), (3d) for recursion], most of the results obtained come in symmetrical pairs, as above, where the proof of the second is analogous to the proof of the first. Only one of the pair of proofs will actually be presented.

We will require below only the following special case of (I) of Lemmas 6, 7 (which, however, could not be proved without the general case).

LEMMA 8. Suppose that $D = ([B]_1, \cdots, z[B]_2)$ is a derivation in G' and that $C \neq B$. Then

(a) if $C \to$ AB, there is a derivation

$$([CB]_1, \cdots, z[C]_2) \qquad \text{in } G'$$

(b) if $C \to BA$, there is a derivation

$$([C]_1, \cdots, z[CB]_2) \qquad \text{in } G'.$$

DEFINITION 10. Suppose that G' is formed from G by the construction and D is an α-derivation of x in G. D will be said to be *represented* in G' if and only if $\alpha = a$ or $\alpha = A$ and there is a derivation $([A]_1, \cdots, x[A]_2)$ in G'.

What we are now trying to prove is that every S-derivation of G is represented in G'.

DEFINITION 11. Let $D_1 = (\varphi_1, \cdots, \varphi_m)$ and $D_2 = (\psi_1, \cdots, \psi_n)$ be derivations in G. Then $D_1{*}D_2$ is the derivation

$$(\varphi_1\psi_1, \varphi_2\psi_1, \cdots, \varphi_m\psi_1, \varphi_m\psi_2, \cdots, \varphi_m\psi_n).$$

LEMMA 9. Let D_1 be an A-derivation of x and D_2 a B-derivation of y in G. If D_1 and D_2 are represented in G' and $C \to AB$, then

$$D_3 = (C\varphi_1 \cdots \varphi_m)$$

is represented in G', where $(\varphi_1, \cdots, \varphi_m) = D_1{*}D_2$. ($D_3$ is thus a C-derivation of xy.)

PROOF. By hypothesis, there are derivations

$$([A]_1, \cdots, x[A]_2) \tag{18}$$

$$([B]_1, \cdots, y[B]_2) \tag{19}$$

in G'.

Case 1. Suppose $A \neq C \neq B$. Then by Lemma 8, there are derivations

$$([C]_1, \cdots, x[CA]_2) \tag{20}$$

$$([CB]_1, \cdots, y[C]_2) \tag{21}$$

in G'. By (iib) of the construction,

$$[CA]_2 \rightarrow [CB]_1 \tag{22}$$

Combining (20), (22), and (21), we have the required derivation.

Case 2. $C = A$. $\therefore C \neq B$ by assumption of regularity of G. By Lemma 8, case (a), we have again the derivation (21). By (iva) of the construction,

$$[A]_2 = [C]_2 \rightarrow [CB]_1 . \tag{23}$$

Combining (18), (23), (21) we have the required derivation.

Case 3. $C = B$. $\therefore C \neq A$. By Lemma 8, case (b), we have (20). By (iiib),

$$[CA]_2 \rightarrow [C]_1 = [B]_1 . \tag{24}$$

Combining (20), (24), (19), we have the required derivation.

Since $C \rightarrow CC$ is ruled out by assumption of regularity, these are the only possible cases.

LEMMA 10. If $D_1 = (\varphi_1, \cdots, \varphi_r)$ is a $\chi_1\omega_1$-derivation, where $\chi_1 \neq I \neq \omega_1$, then there is a derivation $D_2 = D_3{*}D_4 = (\psi_1, \cdots, \psi_r)$ such that $\psi_r = \varphi_r$, D_3 is a χ_1-derivation and D_4 is an ω_1-derivation.

PROOF. Since for $i > 1$, φ_i is formed from φ_{i-1} by replacement of a single symbol of φ_{i-1},[17] we can clearly find χ_i, ω_i s.t. $\varphi_i = \chi_i\omega_i$ where either (a) $\chi_i = \chi_{i-1}$ and $\omega_{i-1} \rightarrow \omega_i$ or (b) $\chi_{i-1} \rightarrow \chi_i$ and $\omega_i = \omega_{i-1}$ ($\chi_{i-1}\omega_{i-1} = \varphi_{i-1}$). Then D_3 is the subsequence of $(\chi_1, \cdots, \chi_r)$ formed by dropping repetitions and D_4 is the subsequence of $(\omega_1, \cdots, \omega_r)$ formed by dropping repetitions.

LEMMA 11. If G' is formed from G by the construction, then every α-derivation D in G is represented in G'.

[17] Which, however, may not be uniquely determined. Compare footnote 8.

PROOF. Obvious, in case D contains 2 lines. Suppose true for all derivations of fewer than r lines ($r > 2$). Let $D = (\varphi_1, \cdots, \varphi_r)$, where $\varphi_1 = \alpha$. Since $r > 2$, $\alpha = A$, $\varphi_2 = BC$. $\therefore$ $(\varphi_2, \cdots, \varphi_r)$ is a BC-derivation. By Lemma 10, there is a $D_2 = D_3{}^*D_4 = (\psi_2, \cdots, \psi_r)$ s.t. D_3 is a B-derivation, D_4 a C-derivation, and $\psi_r = \varphi_r$. By inductive hypothesis, both D_3 and D_4 are represented in G'. By Lemma 9, D is represented in G'.

It remains to show that if $([A]_1, \cdots, x[A]_2)$ is a derivation in G', then there is a derivation $(A, \cdots, x)$ in G.

LEMMA 12.[18] Suppose that G' is formed by the construction from G, regular and non-s.e., and that

(a) $D = (\varphi_1, \cdots, \varphi_{m_1}, \cdots, \varphi_q, \cdots, \varphi_{m_2}, \cdots, \varphi_r)$ is a derivation in G', where $Q_1 = [A_1]$, $Q_{m_1} = Q_{m_2} = [A_1 \cdots A_k]_n$, $Q_q = [A_1 \cdots A_j]_\iota$, $Q_r = [A_1]_2$

(b) there is no u, v s.t. $u \neq v$, $Q_u = Q_v = [B_1 \cdots B_s]_t$, and $s < k$[19]

(c) for $m_1 < u < m_2$, if $Q_u = [A_1 \cdots A_s]_t$, then $s \geqq j$[20]

Then it follows that

(A) if $n = 2$, there is an $m_0 < m_1$ such that $Q_{m_0} = [A_1 \cdots A_k]_1$

(B) if $n = 1$, there is an $m_3 > m_2$ such that $Q_{m_3} = [A_1 \cdots A_k]_2$

(C) $j \geqq k$[21]

PROOF. (A) Suppose $n = 2$. Assume φ_{m_1} to be the earliest line to contain $[A_1 \cdots A_k]_2$. Clearly there is an $\bar{m} \leqq m_1$ s.t. $Q_{\bar{m}} = [A_1 \cdots A_{k+t}]_2$, $Q_{\bar{m}-1} = [A_1 \cdots A_{k+t}]_1 (t \geqq 0)$. If there is no $m_0 < \bar{m}$ s.t.

$$Q_{m_0} = [A_1 \cdots A_k]_1 ,$$

then there must be a $u < \bar{m}$ s.t. $Q_u = [A_1 \cdots A_s]_2$,

$$Q_{u+1} = [A_0 \cdots A_{k-1}B_0 \cdots B_m]_1 ,$$

where φ_{u+1} is formed by (iva) of the construction, $A_0 = I$, $B_0 = A_k$, $m \geqq 1$, and $s < k$ (since (iva) gives the only possibility for increasing the length of Q by more than 1). $\therefore$ $B_{m-1} \rightarrow A_sB_m$. $\therefore$ $A_s \Rightarrow \varphi A_sB_m\psi$.

But $Q_u = [A_1 \cdots A_s]_2$ cannot recur in any line following φ_{m_2} [this would contradict assumption (b)]. Therefore, just as above, there must be a $v > m_2$ s.t. $Q_{v-1} = [A_0 \cdots A_{s-1}C_0 \cdots C_{m'}]_2$, $Q_v = [A_1 \cdots A_p]_1$, where φ_v is formed by (iiib) of the construction, $A_0 = I$, $C_0 = A_s$, $m' \geqq 1$, $p < s$ (since (iiib) gives the only possibility for decreasing the

[18] We continue to employ Convention 2, above.

[19] That is, $Q_{m_1} = Q_{m_2}$ is the shortest Q of D that repeats.

[20] That is, Q_q is the shortest Q of this form between φ_{m_1} and φ_{m_2}.

[21] That is, Q_q is not shorter than $Q_{m_1} = Q_{m_2}$.

length of Q by more than 1). $\therefore$ $C_{m'-1} \rightarrow C_{m'}A_p$. But $A_s \Rightarrow \omega_1 C_{m'-1}\omega_2$ (Lemma 3). $\therefore$ $A_s \Rightarrow \omega_1 C_{m'} A_p \omega_2 \Rightarrow \omega_1 C_{m'} \omega_3 A_s \omega_4 \Rightarrow \omega_1 C_{m'} \omega_3 \varphi A_s B_m \psi \omega_4$. Contra., since G is assumed to be non-s.e.

$\therefore$ there is an $m_0 < \bar{m} \leqq m_1$ s.t. $Q_{m_0} = [A_1 \cdots A_k]_1$

(B) Suppose $n = 1$. Proof is analogous.

(C) (I). Suppose $n = 2$. Suppose $j < k$. Suppose i (in Q_q) is 2. Clearly there must be a $v > m_2$ s.t. either $Q_v = [A_1 \cdots A_j]_2$ [which contradicts assumption (b)] or $Q_{v-1} = [A_0 \cdots A_{j-1} C_0 \cdots C_m]_2$, $Q_v = [A_1 \cdots A_p]_1$, where φ_v is formed by (iiib) of the construction, $A_0 = I$, $C_0 = A_j$, $m \geqq 1$, $p < j$ [as in the second paragraph of the proof of (A)]. Suppose the latter. $\therefore$ $C_{m-1} \rightarrow C_m A_p$. $\therefore$ $A_j \Rightarrow \varphi C_m A_p \psi$. Furthermore, since $p < j$, $A_j \Rightarrow \varphi C_m \omega_1 A_j \omega_2$.

From assumption (c) and assumption of regularity of G, it follows that φ_{q+1} can only have been formed by (iva) of the construction. $\therefore$ $Q_{q+1} = [A_0 \cdots A_{j-1} D_0 \cdots D_t]_1$, where $A_0 = I, D_0 = A_j, t \geqq 1$. $\therefore$ $D_{t-1} \rightarrow A_j D_t$. $\therefore$ $A_j \Rightarrow \omega_3 A_j D_t \omega_4$. $\therefore$ $A_j \Rightarrow \omega_3 \varphi C_m \omega_1 A_j \omega_2 D_t \omega_4$, and A_j is self-embedded, contrary to assumption.

Suppose that i (in Q_q) is 1. By (A), there is an $m_0 < m_1$ s.t. $Q_{m_0} = [A_1 \cdots A_k]_1$. $\therefore$ there is a $u < m_0$ s.t. either $Q_u = [A_1 \cdots A_j]_1$ [which contradicts assumption (b)] or $Q_u = [A_1 \cdots A_s]_2$,

$$Q_{u+1} = [A_0 \cdots A_{j-1} B_0 \cdots B_m]_1,$$

where φ_{u+1} is formed by (iva), $A_0 = I$, $B_0 = A_j$, $m \geqq 1$, $s < j$. Assuming the latter, we conclude that $A_j \Rightarrow \omega_1 A_j \omega_2 B_m \psi$, as above.

From assumption (c) and assumption of regularity of G, it follows that φ_q can only have been formed by (iiib). Contradiction follows as above.

(II) Suppose $n = 1$. Proof is analogous.

This completes the proof. From Lemma 12 it follows readily by the same kind of reasoning as above that

Corollary. Under the assumptions of Lemma 12,

(A) if $n = 2$, φ_{m_1+1} is formed by (iva) of the construction

(B) if $n = 1$, φ_{m_2} is formed by (iiib) of the construction

(C) Q_u is of the form $[A_1 \cdots A_k B_0 \cdots B_s]_t$ ($s \geqq 0, B_0 = I$), for u such that: (a) where $n = 2$ and m_0 is as in (A), Lemma 12, then $m_0 < u < m_2$; (b) where $n = 1$ and m_3 is as in (B), Lemma 12, then $m_1 < u < m_3$.

Furthermore, for $m_1 < u < m_2$, $s > 0$ if $t \neq n$.

Definition 12. Let $D = (\varphi_1, \cdots, \varphi_r)$ be a derivation in G' formed by the construction from G. Then D' *corresponds* to D if D' is a derivation of z_r[22] in G and for each $i, j, k(i < j)$ such that

(a) φ_i is the earliest line containing $[A_1 \cdots A_k]_1$

(b) φ_j is the latest line containing $[A_1 \cdots A_k]_2$

(c) there is no p, q s.t. $i < p < j$, $q < k$, and $Q_p = [A_1 \cdots A_q]_n$,

there is a subsequence $(z_i A_k \psi, \cdots, z_j \psi)$ in D'.

LEMMA 13. Let $D = (\varphi_1, \cdots, \varphi_r)$ be a derivation in G' formed by the construction from a regular, non-s.e. G. Suppose that $Q_1 = [A_1 \cdots A_s]_1$, $Q_r = [A_1 \cdots A_s]_2$, and there is no p, q such that $1 < p < r$, $q < s$, $Q_p = [A_1 \cdots A_q]_n$.

Then there is a derivation $D' = (\varphi_1', \cdots, \varphi_{r'}')$ corresponding to D.

PROOF. Proof is by induction on the number of recurrences of symbols Q_i in D (i.e., the number of cycles in the derivation).

Suppose that there are no recurrences of any Q_i in D. It follows that there can have been no applications of (iva) in the construction of D, i.e., no pairs $Q_i = [A_1 \cdots A_j]_2$, $Q_{i+1} = [A_1 \cdots A_k]_1$ where $j < k$. For suppose there were such a pair. $\therefore$ $A_{k-1} \to A_j A_k$. Also, $j > s$, or Q_i is repeated as Q_r. Clearly there is an $m > i + 1$ s.t. $Q_m = [A_1 \cdots A_{k+n}]_2$ $(n \geqq 0)$. $\therefore$ there is a $t > m$ s.t. either $Q_t = [A_1 \cdots A_j]_2$ (contrary to assumption of no repetitions) or

$$Q_t = [A_1 \cdots A_{j+u}]_2, \quad Q_{t+1} = [A_1 \cdots A_{j-v}]_1 \qquad (u, v \geqq 1),$$

where φ_{t+1} is formed by (iiib). $\therefore$

$$A_{j+u-1} \to A_{j+u} A_{j-v} \Rightarrow A_{j+u}\omega_1 A_{k-1}\omega_2 \to$$
$$A_{j+u}\omega_1 A_j A_k \omega_2 \Rightarrow A_{j+u}\omega_3 A_{j+u-1}\omega_4 A_k \omega_2,$$

contrary to the assumption that G is non-s.e. Similarly, there can be no applications of (iiib) in the construction of D. But now the proof for this case follows immediately by induction on the length of D.

Suppose now that the lemma is true for every derivation containing $<n$ occurrences of repeating Q's. Suppose that D contains n such occurrences.

I.

1. Suppose that the shortest recurring Q in D is $[A_1 \cdots A_k]_2$.

2. Select m_1, m_2 s.t. $m_1 < m_2$; $Q_{m_1} = [A_1 \cdots A_k]_2 = Q_{m_2}$; there is no i, $m_1 < i < m_2$, s.t. $Q_i = [A_1 \cdots A_k]_2$; there is no $j > m_2$ s.t. $Q_j = [A_1 \cdots A_k]_2$.

[22] Compare Convention 2.

3. By Lemma 12 (A), we know that there is an $m_0 < m_1$ s.t. $Q_{m_0} = [A_1 \cdots A_k]_1$. Select m_0 as the earliest such (there is in fact only one). By the Corollary to Lemma 12, (C), and the inductive hypothesis, there is a derivation $D_1 = (z_{m_0}A_k, \cdots, z_{m_1})$[23] corresponding to $(\varphi_{m_0}, \cdots, \varphi_{m_1})$.

4. By Corollary (A), we know that

$$\varphi_{m_1+1} = z_{m_1}[A_0 \cdots A_{k-1}B_0 \cdots B_m]_1,$$

where $A_0 = I$, $B_0 = A_k$, $m \geqq 1$, $B_{m-1} \rightarrow A_kB_m$. Obviously, there is a $v(m_1 < v < m_2)$ s.t. either $Q_v = [A_0 \cdots A_{k-1}B_0 \cdots B_m]_2$ or

$$Q_v = [A_0 \cdots A_{k-1}B_0 \cdots B_{m+t}]_2, \; Q_{v+1} = [A_0 \cdots A_{k-1}B_0 \cdots B_{m-u}]_1,$$

where $u, t > 1$ and φ_{v+1} is formed by (iiib) [note Corollary (C)]. From the latter assumption we can deduce self-embedding, as above. $\therefore$ we can select v as the largest integer $< m_2$ s.t. $Q_v = [A_0 \cdots A_{k-1}B_0 \cdots B_m]_2$.

5. Let t be the largest integer $(m_1 + 1 < t < v)$ s.t.

$$Q_t = [A_0 \cdots A_{k-1}B_0 \cdots B_{m-u}]_i, \qquad u > 0.$$

Suppose that $i = 1$. But φ_{t+1} must be formed by (iia) or (iiia) of the construction. $\therefore$ $u = 1$ and $B_{m-1} \rightarrow B_mC$, contrary to assumption of regularity, since $B_{m-1} \rightarrow A_kB_m$.

$\therefore$ $i = 2$, and $Q_{t+1} = [A_0 \cdots A_{k-1}B_0 \cdots B_{m+n}]_1 (n \geqq 0)$, where φ_{t+1} is formed by (iva) of the construction. $\therefore$

$$B_{m+n-1} \rightarrow B_{m-u}B_{m+n} \Rightarrow \omega_1 B_{m-1}\omega_2 B_{m+n} \rightarrow \omega_1 A_k B_m \omega_2 B_{m+n}.$$

Suppose $n = 0$. Then $B_{m-1} \rightarrow B_{m-u}B_m$, so that, by regularity, $B_{m-u} = A_k$. $\therefore$ $Q_t = [A_1 \cdots A_k]_2$, contrary to assumption in step 2.

$\therefore$ $n > 0$. $\therefore$ $B_m \Rightarrow \omega_3 B_{m+n-1}\omega_4$. $\therefore$ $B_m \Rightarrow \omega_3\omega_1 A_k B_m \omega_2 B_{m+n}\omega_4$, contra. (s.e.).

6. $\therefore$ there is no t such as that postulated in step 5. Consequently $(\varphi_{m_1+1}, \cdots, \varphi_v)$ meets the assumption of the inductive hypothesis[24] and there is a derivation $D_2 = (z_{m_1+1}B_m, \cdots, z_{m_v})$[25] corresponding to $(\varphi_{m_1+1}, \cdots, \varphi_v)$.

7. Since v was selected in step 4 to be maximal, it follows that φ_{v+1} cannot be formed by (iva), by reasoning similar to that involved in

[23] Recall that $z_{m_1} = z_{m_0}z_{m_1}^{m0+1}$; i.e., there is a derivation $(A_k, \cdots, z_{m_1}^{m0+1})$.

[24] From nonexistence of such a t it follows at once that for u such that $m_1 < u < v$, $Q_u = [A_0 \cdots A_{k-1}B_0 \cdots B_mC_0 \cdots C_{\bar{m}}]_i$ $(\bar{m} \geqq 0, C_0 = I)$.

[25] That is, there is a derivation $(B_m, \cdots, z_{m_v}^{m_1+2})$.

step 4. By regularity assumption, it cannot have been formed by (iib) or (iiib) of the construction, since $B_{m-1} \to A_k B_m$. $\therefore$

$$Q_{v+1} = [A_0 \cdots A_{k-1} B_0 \cdots B_{m-1}]_2 .$$

8. Suppose $m = 1$, so that $Q_{v+1} = [A_1 \cdots A_k]_2$. $\therefore$ $v + 1 = m_2$, by assumption of step 2, and $A_k \to A_k B_1$. Let D_2' be the derivation formed from D_2 (cf. step 6) by deleting initial z_{m_1+1} from each line. Let

$$(\psi_1 , \cdots, \psi_p) = D_1{}^* D_2'$$

(cf. Definition 11; D_1 as in step 3). Clearly $D_3 = (z_{m_0} A_k , \psi_1 , \cdots, \psi_p)$ is a derivation corresponding to $(\varphi_{m_0} , \cdots, \varphi_{m_2})$.

9. Suppose $m \geqq 2$. By assumption that G is non-s.e., and that v is maximal (in step 4) we can show that φ_{v+2} must be formed by (iib) of the construction (all other cases lead to contradiction). $\therefore$

$$Q_{v+2} = [A_0 \cdots A_{k-1} B_0 \cdots B_{m-2} C]_1 , \quad B_{m-2} \to B_{m-1} C.$$

As above, we can find a v_1 which is the largest integer $< m_2$ s.t. $Q_{v_1} = [A_0 \cdots A_{k-1} B_0 \cdots B_{m-2} C]_2$ and s.t. $(\varphi_{v+2} , \cdots, \varphi_{v_1})$ meets the inductive hypothesis. $\therefore$ there is a derivation $D_4 = (z_{v+2} C, \cdots, z_{v_1})$ corresponding to $(\varphi_{v+2} , \cdots, \varphi_{v_1})$.

10. Suppose $m = 2$. $\therefore$

$$B_{m-2} \to B_1 C, \quad v_1 + 1 = m_2 , \quad Q_{v_1+1} = [A_1 \cdots A_k]_2$$

(as above). Let D_4' be the derivation formed from D_4 by deleting initial z_{v+2} from each line. Let $(\psi_1 , \cdots, \psi_p)$ be as in step 8. Let $(\chi_1 , \cdots, \chi_q) = (z_{m_0} B_1 , \psi_1 , \cdots, \psi_p)^* D_4'$. Clearly $D_5 = (z_{m_0} A_k , \chi_1 , \cdots, \chi_q)$ is a derivation corresponding to $(\varphi_{m_0} , \cdots, \varphi_{m_2})$.

11. Similarly, whatever m is, we can find a derivation

$$\Delta = (z_{m_0} A_k , \cdots, z_{m_2})$$

corresponding to $(\varphi_{m_0} , \cdots, \varphi_{m_2})$.

12. Consider now the derivation D_6 formed by deleting from the original D the lines $\varphi_{m_1+1} , \cdots, \varphi_{m_2}$ and the medial segment $z_{m_2}^{m_1+1}$ from each later line. That is, $D_6 = (\psi_1 , \cdots, \psi_t)$ $(t = r - (m_2 - m_1))$, where for $i \leqq m_1$, $\psi_i = \varphi_i$, and for $i > m_1$, $\psi_i = z_{m_1} z_{m_2-m_1+i}^{m_2+1} Q_{m_2-m_1+i}$. By inductive hypothesis, there is a derivation D_7 corresponding to D_6 .

In steps 2 and 3, m_0 , m_1 , m_2 were chosen so that φ_{m_0} contains the earliest occurrence of $Q_{m_0} = [A_1 \cdots A_k]_1$, and φ_{m_2} the latest occurrence of $Q_{m_2} = Q_{m_1} = [A_1 \cdots A_k]_2$, and so that no occurrences of Q_{m_1} occur between φ_{m_1} and φ_{m_2}. $\therefore$ in D_6, ψ_{m_0} contains the earliest occurrence of Q_{m_0} and ψ_{m_1} the latest occurrence of Q_{m_1}. Furthermore, by Corollary (C), there is no Q shorter than Q_{m_0} between φ_{m_0} and φ_{m_1}. $\therefore$ by inductive hypothesis and the definition of correspondence, it follows that D_7 contains a subsequence $\bar{D}_7 = (z_{m_0}A_k\bar{\psi}, \cdots, z_{m_1}\bar{\psi})$. But step 11 guarantees us a derivation $\Delta = (z_{m_0}A_k, \cdots, z_{m_2})$ corresponding to $(\varphi_{m_0}, \cdots, \varphi_{m_2})$. We now construct D_8 by replacing $\bar{D}_7$ in D_7 by $\bar{\Delta} = (z_{m_0}A_k\bar{\psi}, \cdots, z_{m_2}\bar{\psi})$, formed by suffixing $\bar{\psi}$ to each line of Δ, and inserting $z_{m_2}^{m_0+1}$ after z_{m_0} in all lines of D_7 following the subsequence $\bar{D}_7$.

Clearly D_8 corresponds to D, which is the required result in case the shortest recurring Q is of the form $[\cdots]_2$.

II.

An analogous proof can be given for the case in which the shortest recurring Q is of the form $[\cdots]_1$.

We have shown that the lemma holds for derivations with no recursions, and that it holds of a derivation with n occurrences of recurring Q's on the assumption that it holds for all derivations with $<n$ such occurrences. $\therefore$ it is true of all derivations.

A corollary follows immediately.

Corollary. If G' is formed from G by the construction and $D' = ([A]_1, \cdots, x[A]_2)$ is a derivation in G', then there is a derivation $D = (A, \cdots, x)$ in G.

From this result and Lemma 11, we draw the following conclusion.

Theorem 10. If G' is formed from G by the construction, then there is a derivation $(S, \cdots, z)$ in G if and only if there is a derivation $([S]_1, \cdots, z[S]_2)$ in G'.

That is, if $[S]_1$ in G' plays the role of S in G, then G and G' are equivalent if we emend the construction by adding the rule $Q_1 \rightarrow a$ wherever there are $Q_2, \cdots, Q_n$ $(n \geqq 2)$ such that $Q_1 \rightarrow aQ_2$ and $Q_2 \rightarrow Q_3 \rightarrow \cdots \rightarrow Q_n$, where $Q_n = [S]_2$, Q_i is of the form $[\cdots]_2$ for $1 < i \leqq n$, and Q_1 is of the form $[\cdots]_1$.

But in the grammar thus formed all rules are of the form $A \rightarrow aB$ (where a is I unless the rule was formed by step (i) of the construction) or $A \rightarrow a$. It is thus a type 3 grammar, and the language L_G generated by G could have been generated by a finite state Markov source (cf. Theorem 6) with many vacuous transitions. But for every such source, there is an equivalent source with no identity transitions (cf. Chomsky and Miller, 1958). Therefore L_G could have been generated by a finite Markov source of the usual type. Obviously, every type 3 grammar is non-s.e. (the lines of its A-derivations are all of the form xB). Consequently:

THEOREM 11. If L is a type 2 language, then it is not a type 3 (finite state) language if and only if all of its grammars are self-embedding.

Among the many open questions in this area, it seems particularly important to try to arrive at some characterization of the languages of these[26] various types[27] and of the languages that belong to one type but not the next lower type in the classification. In particular, it would be interesting to determine a necessary and sufficient structural property that marks languages as being of type 2 but not type 3. Even given Theorem 11, it does not appear easy to arrive at such a structural characterization theorem for those type 2 languages that are beyond the bounds of type 3 description.

RECEIVED: October 28, 1958.

REFERENCES

CHOMSKY, N. (1956). Three models for the description of language. *IRE Trans. on Inform. Theory* **IT-2,** No. 3, 113–124.

CHOMSKY, N. (1957). "Syntactic Structures." Mouton and Co., The Hague.

CHOMSKY, N., and MILLER, G. A. (1958). Finite state languages. *Inform. and Control* **1,** 91–112.

DAVIS, M. (1958). "Computability and Unsolvability." McGraw-Hill, New York.

HARRIS, Z. S. (1952a). Discourse analysis. *Language* **28,** 1–30.

HARRIS, Z. S. (1952b). Discourse analysis: A sample text. *Language* **28,** 474–494.

HARRIS, Z. S. (1957). Cooccurrence and transformation in linguistic structure. *Language* **33,** 283–340.

KLEENE, S. C. (1956). Representation of events in nerve nets. *In* "Automata Studies" (C. E. Shannon and J. McCarthy, eds.), pp. 3–40. Princeton Univ. Press, Princeton, New Jersey.

POST, E. L. (1944). Recursively enumerable sets of positive integers and their decision problems. *Bull. Am. Math. Soc.* **50,** 284–316.

SCHÜTZENBERGER, M. P. (1957). Unpublished paper.

[26] And several other types. In particular, investigations of this kind will be of limited significance for natural languages until the results are extended to transformational grammars. This is a task of considerable difficulty for which investigations of the type presented here are a necessary prerequisite.

[27] As, for example, the results cited in footnote 4 characterize finite state languages.