ESSAYS IN LINGUISTICS

By

JOSEPH H. GREENBERG

Essays in Linguistics

By

JOSEPH H. GREENBERG

THE UNIVERSITY OF CHICAGO PRESS

This volume has also been issued by the Wenner-Gren Foundation for Anthropological Research, Incorporated, as *Viking Fund Publications in Anthropology Number 24*, in a limited, paper-bound edition for private distribution to scholars and institutions throughout the world. The publishers gratefully acknowledge the permission granted by the Foundation for the appearance of this edition.

Library of Congress Catalog Number: 57-6273

THE UNIVERSITY OF CHICAGO PRESS, CHICAGO 37
Cambridge University Press, London, N.W. 1, England
The University of Toronto Press, Toronto 5, Canada

Published 1957. Second Impression 1958.

Essays in Linguistics

PREFACE

The essays of this collection are intended as separate treatments of a number of topics in linguistics. They fall quite naturally into three groups, the first two being concerned with the methodology of language description, the third and fourth with historical linguistics, and the remaining four with the relation between language and culture. Though not designed to cover the entire field of linguistics, almost every topic comes in for some discussion. The most serious omissions are the absence of any treatment of phonemic theory and of any over-all discussion of meaning, though semantic problems are touched on incidentally at several points.

There are obvious interconnections among the ideas expressed in some of the essays, though they are not meant to present any coherent "system." In the nature of things, problems as diverse as those dealt with here often have solutions which do not depend on one another. If there is any single point of view that runs through the whole, it is that further substantial progress in linguistics requires the abandonment of its traditional isolationism, one for which there was formerly much justification, in favor of a willingness to explore connections in other directions. The borderline areas most prominent in the present essays are those with logic, mathematics, anthropology, and psychology, but, of course, others exist.

I have written chiefly for those anthropologists, in whatever branch of the subject they are engaged, who, because of their interest in cultural theory, are aware of the significance of so fundamental a human trait as language to any general science of man. The essays are reasonably independent of one another, and the less linguistically oriented anthropologist who finds the first two essays in particular somewhat technical and remote from his main interests should have no compunction in passing them by. On the other hand, these may well be precisely the essays which hold the most real interest for the logician or mathematician interested in the possibility of a general syntax, of which linguistics would be but a branch. The mathematical reader should bear in mind that my own command of mathematics is very far from professional, and, in view of the purposes of the essays as a whole, I have not assumed any but an elementary acquaintance with mathematics on the part of the reader. For this reason also, a few topics of more purely mathematical interest have been relegated to appendixes.

None of the essays has appeared elsewhere in its present form. However, the essay "Genetic Relationship among Languages" is an expanded and revised version of "Historical Linguistics and Unwritten Languages," which appeared in *Anthropology Today*, edited by A. L. Kroeber (Chicago, 1952), pages 265–87; and parts of

the discussion of the definition of the word and the morpheme occur in an article entitled "A Quantitative Approach to the Morphological Typology of Language," in *Methods and Perspectives in Anthropology: Papers in Honor of Wilson D. Wallis*, edited by Robert F. Spencer (Minneapolis, 1954), and in *Psycholinguistics: A Survey of Theory and Research Problems* (Baltimore, 1954). "Language and Evolutionary Theory" was the topic of a talk at the Wenner-Gren Foundation in 1951, and the subject matters of "Order of Affixing" and "Language as a Sign System" were discussed in talks at Michigan and Yale, respectively, in 1954. On all these occasions, I derived benefit from oral comments and criticisms.

I am grateful to the Ford Foundation, which provided the necessary leisure, under a Faculty Research Fellowship, to pursue my interest in logic and mathematics and to write the essays as a whole. I am also in the debt of the Social Science Research Council for the opportunity to participate in its summer seminar of 1953 in psycholinguistics at the University of Indiana, at which some of the ideas regarding the definition of the word were first developed and presented in oral discussion and which also stimulated my interest in the connection between language universals and general psychology, which figures in the final essay.

The first version of these essays was written during the summer of 1954, and no systematic account is taken of the literature which has appeared since that time.

I am indebted to Professor Marvin Harris for observations regarding the difference between scientific and ordinary language, which helped to orient my thinking in the area of language and evolution, and to Professors Charles Osgood and Floyd Lounsbury, whose discussion of psychological learning theory in relation to prefixing and suffixing provided the initial stimulus for the essay on the order of affixed elements.

Finally, and by no means least, I owe a debt of gratitude to my wife Selma for her sympathetic understanding during the period I was writing these essays.

JOSEPH H. GREENBERG

TABLE OF CONTENTS

CHAPTER I

LANGUAGE AS A SIGN SYSTEM

LANGUAGE can be approached in either of two ways: as a system of signals conforming to the rules which constitute its grammar or as a set of culturally transmitted behavior patterns shared by a group of individuals. It is the first of these aspects that will interest us for the present; but in later chapters of this work language as the heritage of cultural groups will receive a major share of attention.

As a system, it is natural to compare spoken language with other forms which likewise consist of ordered arrays of elements in some physical medium and obey fixed rules of combination. For example, the expressions of mathematics seem to display a certain analogy with language. From a limited number of elementary symbols, sequences of finite length are built in conformity with certain rules which may be called the "grammar" of the system. A general discipline whose subject matter embraces such systems would contain the study of linguistic systems as a subdivision. In this way the analytical problems of language might be put into the broader perspective of a semiotic whose terminology would apply equally to linguistic and to non-linguistic systems.

Such a branch of inquiry belongs to the logico-mathematical group of studies, since it will consider the nature of all possible systems with postulated characteristics. Even its linguistic subdivision will be distinct from, though in intimate relation with, linguistics, which remains an empirical social science dealing with the description, history, and ethnolinguistics of actually existing languages.

Our first task will be to establish more precisely than has been possible in these few preliminary remarks the defining characteristics of the systems to be included in such a general semiotic. The well-known distinction among three aspects of sign behavior—the syntactic, the semantic, and the pragmatic—introduced by Charles W. Morris will serve as a convenient point of departure.[1] This analysis starts from the sign situation as involving three factors—the sign itself, the referent, and the organism who produces or reacts to the sign. The syntactic aspect is one in which only the relation of sign to sign is included, abstractions being made of both the referent and the organism. In the semantic aspect the relation between sign and referent is taken into account, but the organism is excluded. In the pragmatic aspect, all three—the organism, the sign, and the referent—are taken into consideration. The pragmatic aspect is usually understood as treating of the actual behavior of organisms in their use of the sign system as opposed to the rules of the sign system

1. In his *Foundations of the Theory of Signs* (*International Encyclopedia of Unified Science*, Vol. I, No. 2 [Chicago, 1935]).

viewed as a calculus without regard to meaning, which is the subject matter of syntactics, and to the meanings which belong to semantics.

If we direct our attention to semantics for a moment, however, it can be seen that a distinction can be made within semantics analogous to that between syntactics and pragmatics. We have, on the one hand, semantic rules, rules of meaning, and, on the other hand, the actual behavior of organisms in regard to meaning in their use of the language. The former is a kind of syntactic-semantics, the latter a pragmatic-semantics. If we turn now to the sign as a physical object, a similar differentiation can be made between the rule which specifies which physical phenomena shall be instances of a particular sign vehicle and the actual behavior of sign-using organisms in this regard. The same distinction also holds between rules of sign arrangement and actual behavior in regard to sign arrangement. To avoid confusion with the technical use of the term "syntax" in language, we may call "systemic" those investigations having to do with the formulation and discovery of rules and retain the term "pragmatic" to refer to the behavior of organisms in their use of systems. To designate rules concerning the arrangements of signs without regard to meaning, the term "grammar" may be extended from its employment in linguistics to cover sign systems in general. This results in six aspects of sign systems, as indicated in the accompanying table. Then linguistics is concerned with

	Systemic	Pragmatic
Physical		
Semantic		
Grammatical		

the systemic aspects of a particular group of actually existing sign systems, the so-called "natural languages," while psychology and the social sciences, in so far as they deal with verbal behavior, consider the pragmatic aspects of these same systems. Although it was not drawn up with this purpose in view, the present analysis tends to justify the traditional division of language descriptions into phonology (the physical aspect), lexicography (the semantic aspect), and grammar (the grammatical aspect).

Confining ourselves henceforth to the realm of the systemic, we can see that, of the three aspects—the physical, the semantic, and the grammatical—not all are equally indispensable. It is possible to have systems of elements subject to certain specified rules of arrangements but without any meanings assigned to the elements. Such a system will have physical and grammatical, but no semantic, rules. A system which lacks semantic rules may be called a "calculus" or an "uninterpreted system." Systems which include rules of meaning will be called "interpreted systems."

A further distinction may be made between systems, here called "specified," in which the nature of the physical objects which are to function as the actual signs is specified by rules, and abstract systems, in which it is not.[2] The physical aspect

2. The use of the term "abstract" here corresponds to its employment in group theory where groups with different elements but the same structure, and hence connected by one-to-one isomorphisms, are said to represent the same abstract group.

is thus also dispensable. The various specified systems which are realizations of the same abstract system are connected by a relation of isomorphism. A common example of isomorphic systems is a spoken language and its written form in phonemic transcription. For an isomorphism to exist, we must have a set of one-to-one transformations of the two systems which carries every expression of one language into its corresponding expression in the other. The monoalphabetic substitution ciphers of the Sunday supplement cryptograms are examples of systems isomorphically related to written English by element-for-element substitution rules.

The grammatical aspect alone is indispensable. There can be no system without rules of arrangement. As an inclusive designation for all systems, whether calculi or interpreted systems, whether specified or abstract, the term "sign system" will be used (hereafter abbreviated to "SS").[3] As a logical minimum for a sign system, we require a set of elements, whether specified or not, ordered into sequences called "expressions" by a serial relation and conforming to definite rules of combination. The number of elements may be finite or infinite.[4] The number of expressions may likewise be finite or infinite. A system will be called "finite" or "infinite," respectively, depending on whether the number of its expressions is finite or infinite. The basic serial relation of a sign system, if specified, must be defined along with the physical shape of its elements. We take the ordering in time of the elements of spoken language so for granted that this tends to be forgotten. As soon as we specify an isomorphic system in the visual medium of writing, we see that a direction, whether from left to right, from right to left, or downward, etc., must be defined as the isomorphic substitute of the relation "following in time." In addition to the basic serial relation, a system may have others. In language, sentence intonations, which are in relation to an expression as a whole, furnish an example. In mathematics the relation between a numeral on the line and a power written above and to the right is an additional relation. Thus 34 and 3^4 contain the same elements connected by different relations. A more rigorous statement of the requirement of a single serial relation, as well as a discussion of types of systems generated by additional relations, is to be found in Appendix I, "On Basic Relations in Sign Systems."

Two specified systems whose elements are identical if equal in number, or such that all the elements of the system with the smaller number of elements are identical with some elements in the larger system, are said to be "homogeneous"; otherwise, "heterogeneous." Written English and written French are homogeneous. Written English and written Russian are heterogeneous. The element order of a system

3. This term is inappropriate, since a sign is generally thought of as having some meaning, but no term that will express exactly what is wanted seems available.

4. An example of a system with an infinite number of elements would be one in which the first sign consisted of one vertical line to which were added, perpendicularly on the right, up to, say, five horizontal lines. When this was reached, the next sign would add a vertical line perpendicular to the preceding five horizontal lines. To it, in turn, would be added up to five vertical lines, and so on. A construction in this SS might be limited to a sequence of any two signs of the system. In this case there would be an infinite number of expressions in the system, even though two was the maximum length of any expression.

is the number of elements it contains. Written English has the element order 26. Sometimes, when there is no risk of confusion with the term "sequence order," to be introduced later, this will be called merely "order."

Certain concepts and notations drawn from the mathematical theory of sets or aggregates will prove useful.[5] An SS will be considered a set whose members are the expressions of the system. One SS will be said to be equal to another if both contain the same expressions, in symbols, $M_1 = M_2$. If all the expressions of M_1 are also expressions of M_2 but not all the expressions of M_2 are expressions of M_1, then M_1 is contained in, or is a proper part of, M_2 ($M_1 \subset M_2$). If all the expressions of M_1 are expressions of M_2, then M_1 is equal to, or is contained in, M_2 ($M_1 \subseteq M_2$). The system which contains all the expressions of M_1 and all the expressions of M_2, including those found in both, is called the "union" of M_1 and M_2 ($M_1 \cup M_2$). The system which contains only those expressions which are in both M_1 and M_2 is the intersection of these two systems, $M_1 \cap M_2$. For all homogeneous systems of the same order there is one SS in which all the others are contained. This SS, indicated by I^n, where n indicates the order, is simply the unrestricted set of permutations and combinations of the n symbols and will be called the "infinite system" of that order.

For example, I^4, with symbols specified as *a*, *b*, *c*, and *d*, has four expressions of length (hereafter abbreviated *l*) 1: *a*, *b*, *c*, *d*, and sixteen expressions for $l = 2$: *aa*, *ab*, *ac*, *ad*, *ba*, *bb*, *bc*, *bd*, *ca*, *cb*, *cc*, *cd*, *da*, *db*, *dc*, *dd*. In fact, the number of expressions in I^n of length *l* is l^n. The unrestricted set of permutations and combinations in an SS with an infinite number of elements, I^∞, contains every homogeneous SS of whatever order. In general, if $m < n$, then I^n contains every homogeneous SS of order m.

For every order we have an ideal construction, the null system, which contains no expressions and which corresponds to the empty set. It will be indicated by 0^n. $M_1 - M_2$ indicates the SS which has all those expressions which are in M_1 but not in M_2. To every system M_1, there corresponds another system $-M_1$, called its "complement," consisting only of those homogeneous expressions of the same order which are not members of M_1. This system may be defined as $I - M_1$.

The isomorphism of the set of all homogeneous SS of order $\leq n$ to the algebra of classes is obvious. In fact, it forms a Boolean algebra of an infinite number of elements.[6]

If an SS is finite, i.e., contains a finite number of expressions, then there must be an expression or expressions of maximal length. For example, if an SS is of order 6 and no expression is of length greater than 10, then, even if all combinations of length ≤ 10 are allowed, the maximum number of expressions is 6^{10}, which is

5. Some of the main expositions of set theory are: Erich Kamke, *Theory of Sets*, translated from the 2d German ed. (New York, 1950); Adolph Fraenkel, *Einleitung in die Mengenlehre* (Berlin, 1923); Felix Hausdorff, *Mengenlehre* (3d ed.; Berlin and Leipzig, 1935).

6. For Boolean algebras see the standard works on symbolic logic, particularly Paul C. Rosenbloom, *The Elements of Mathematical Logic* (New York, 1950).

finite. A language of infinite element order, however, must have an infinite number of expressions, even if there are expressions of maximum length. With regard to finiteness, then, there are four classes of systems: (1) those of finite order with expressions of a maximum length *l;* (2) those of finite order without maximum *l;* (3) those of infinite order with maximum *l;* and (4) those of infinite order without maximum *l*. Of these, members of the first class of systems have a finite number of expressions, the others an infinite number. In what follows, except where a statement is made to the contrary, we shall be concerned with systems of finite order and no maximum *l*. This is the class to which all natural languages belong.

Since any SS is contained in the homogeneous infinite SS of the same order (the system containing all permutations and combinations of the elements), the enunciation of grammatical rules is, in essence, the laying of bounds on this infinite system by setting limitations to the allowable permutations and combinations. A rule or set of rules is said to be "well determined" if it is sufficient to allow us to decide for any permutation and combination of its elements, i.e., for any member of the appropriate infinite system, whether the expression belongs to the system or not. The notion of "well-determination" is therefore a test of the adequacy of grammatical rules.

Rules are of a number of possible kinds. The following enumeration is not logically exhaustive.

I. *Cardinal rules.*—These rules have to do with the cardinal number of occurrences of a particular element in the expressions of an SS. They include the following: (1) rules of maxima and minima state that a given element may occur at the most n times or at the least n times in every expression; (2) rules of ratio specify that if two elements, x and y, appear in an expression, x/y is a constant; (3) rules of relative size state that if two elements, x and y, occur in an expression, the number of occurrences of y is $x + n$, where x is a constant.

Example: In an SS of order 5 with the elements specified as a, b, c, d, and e, if we have a rule that the maximum of c is 2, then, according to this well-determined rule, the expression *abcacd* is in the system, but *accdca* is not.

II. *Rules of transition.*—These are rules regarding the limitations on the occurrence of certain elements in certain positions if certain others are found in an expression. They are divided into positive and negative rules, depending on whether the element is required or excluded, and definite or indefinite, depending on whether the relative position of the required or excluded element is defined or not. An example of a positive definite transitional rule is the following: c must always be preceded by a with one other element intervening. An example of a negative indefinite rule is that d may not be preceded at any distance by b. In mathematics and written English, the requirement that an opening parenthesis must always be followed by a closing parenthesis at some interval is a rule of positive indefinite transition.

III. *Rules of infinite interpolation.*—First the elements are divided into two or more classes, which may, but need not, overlap. Every expression consists of some specified number of members of one or more, but not all, the classes and any num-

ber of occurrences (including zero) of members of the other classes. For example, we divide the elements *a*, *b*, *c*, *d*, *e*, into two classes, A, containing *a* and *b*, and B, containing *c*, *d*, and *e*. We then specify that every expression must contain one instance of a member of A and any number of instances of members of B. We might, as an additional rule, also require that the members of B always follow. On this basis, *bded*, *acc*, *b*, are in the system, but *de*, *cab*, *ba*, and *dca* are not. Rules of infinite interpolation are the model for linguistic rules of phrase expansion. They may be considered as rules of transition in which we operate not with individual elements but with classes of elements.

IV. *Rules of length.*—These rules exclude expressions of certain lengths, for example, all those expressions whose length is even. In fact, any monotonically increasing function whose domain is the entire set of positive integers and whose range is included in the positive integers will do, e.g., the function which assigns to each positive integral number n the nth prime number. In this case, all expressions whose length is a prime number are ungrammatical.

V. *Ordinal rules.*—We may make use of the functions just mentioned and apply them not to the length of the expressions but to an ordered set of all the expressions of Rule I. If we assign an order to each element, say, the alphabetic order, *a*, *b*, *c*, *d*, *e*, then we can first list expressions of length 1, then those of length 2, etc., and, within each, follow the dictionary rules of order: *a*, *b*, *c*, *d*, *e*, *aa*, *ab*, *ac*, *ad*, *ae*, *ba*, *bb*, *bc*, *bd*, *be*, *ca*, *cb*, *cc*, *cd*, *ce*, *da*, *db*, *dc*, *dd*, *de*, *ea*, *eb*, *ec*, *ed*, *ee*, *aaa*, *aab*, etc. Then all the values of some function as just described can be included in the system. If, in this case, the function is $y = 2x$ for positive integral x, then the second, fourth, sixth, etc., of the above expressions are grammatical, i.e., in the system, and the first, third, etc., are not.[7]

Instead of using elements as units in applying the foregoing rules, we can use specified finite sequences. For example, in an SS of element order 4 with the elements specified as *a*, *b*, *c*, *d*, we might form the following six sequences: *acdd*, *ba*, *ad*, *cc*, *a*, *bcdcd*. The number of such sequences is the sequence order of the SS, in this case six. I_6 with these specifications will consist of all possible permutations and combinations of these elements.[8] The same kinds of rules can be applied to sequences as to elements. For example, in the present case we can lay down a maximal rule that *ba* may not occur more than once in an expression and a length rule that all expressions contain an odd number of sequences. Then *acddcca* will be in the system, but *baadba* will not because, though of length 3, it contains *ba* twice. Likewise, *acddab* will not be in the system because it is not composed exclusively of the defined sequences. Systems without such sequences, like those of the earlier examples, may be considered limiting cases in which the element order and the sequence order are

7. Actually, all rules can be stated as ordinal rules if we apply the term "function" in a broad manner, as is usual in modern mathematics. The type described in the text under ordinal rules is a special case in which this method proves simplest. For further treatment of this topic see Appendix II, "On Ordinal Rules."

8. The convention is employed of uniting the element order at the upper right and the sequence order at the lower right.

the same and each sequence contains a single element. Since there will frequently be occasion to make statements referring equally to elements or sequences, it will be convenient to have a term "unit" to cover both cases. Similarly, in a system with sequences which are distinct from units, a given expression will contain an equal or greater number of sequences than of elements. To deal with this eventuality the terms "element length," "sequence length," and "unit length" will be employed in analogous fashion.

Various of the foregoing types of rules can be combined as simultaneous requirements. For example, we can require that *c* may not occur more than twice, that *a* never be immediately followed by *b*, and that all expressions have a length which is a multiple of 3. Then *ccc* will not be in the system, because, although its length is a multiple of 3, it contains more than 2 occurrences of *c*. Nor will *cabdec* belong because, although its length is 6, a multiple of 3, and it does not contain more than 2 occurrences of *c*, it exhibits the forbidden sequence *ab*. It is clear that a system governed by a number of rules simultaneously contains only those expressions found in every one of the systems specified by each rule in isolation. It is therefore the intersection of these systems.

Such rules may apply to sign systems both with a finite and with an infinite number of expressions. Finite systems, however, need not conform to any of these rules but may consist of any arbitrary selection of the units. A listing of the allowed combinations is, in this case, a well-determined procedure, since any combination occurring in the list belongs to the system and any which is not found there is excluded. It is obvious that the procedure of listing is always open to us for finite systems but that rules, where possible, will be more convenient. There are also infinite systems defined as $I^n - F^n$, where F^n is some finite system of order n and I and F are homogeneous. Such systems may be defined by a negative list procedure if F^n cannot be defined except by list, for $I^n - F^n$ will consist of all those expressions not listed as belonging to F^n.

A semiotic must contain two distinct classes of procedures. By one, which may be called "synthetic," well-determined rules are stated, in accordance with which the expressions of a given SS may be constructed. The opposite procedure—analysis—which has not been considered in the discussion up to this point, is of particular interest to a science such as linguistics which operates with empirically given systems. The problem here is, given samples of expressions in the system, to derive an adequate set of rules. This involves many additional considerations. Initially, it may be pointed out, equal systems, that is, systems which contain exactly the same expressions, may be defined by different sets of rules. Such systems may be called "heteronomic." The following is a trivial example of heteronomy. The system of element order n (and sequence order n) with the length rule that all expressions must be of even length is equal to the system of element order n and sequence order n^2 without restriction on the combination of sequences. For $n = 3$ and specified elements *a*, *b*, *c*, with the former set of rules the expression *bcbacb* is interpreted as of length 6 composed of the elements *a*, *b*, and *c*, while with the latter it is interpreted as of length 3

and composed of the sequences *bc*, *ba*, and *cb*. We ask whether, among the various ways of formulating rules, there is one which is non-arbitrary. This is equivalent to asking for a procedure which, applied in every case, will pick out one solution among all the possible ones.

Analysis, moreover, raises the question of induction in acute form. Since our sample is always finite, if we have to do with an infinite system, we never know but that the next example will overthrow one of our rules. A form of quantitative inductive logic which assigns a degree of confirmation to each rule, such as that proposed by Carnap, may prove useful here.[9] To take a simple example, if a system has a rule that some element may not occur more than twice in any expression, this rule is better confirmed in one sample than in another if the sample contains a larger number of expressions, if the expressions are longer, and if the system has a smaller number of elements. In the latter instance the proportion of expressions which might break the rule is greater; hence the lack of negative instances in a given sample is a more powerful confirmation of the rule.

Two alternative methodologies are to be considered in testing rules. In the first of these we confine ourselves to a sample in which all the expressions have been spontaneously offered or elicited from the informant. We can then say, concerning any possible expression, whether it belongs to this corpus or not. If it does, it is an expression in the system. If it does not, we cannot reach a decision. We can merely wait to see whether it will occur or be elicited later. On this procedure we can sometimes say that an expression is in the system, but we can never definitively say that it is not.

The second method, which may be called that of the leading question, allows us greater latitude. On this procedure, we are permitted to make up any expression whatever and ask the informant whether it is in the system or not. As contrasted with the first method, it will sometimes allow us to say that an expression is not in the system. It provides a means of swift refutation of a rule in some instances, though not of confirmation. To illustrate, in the case of an element which may not occur more than twice, we make up an expression in which the element occurs three times, thus violating the rule. If the informant accepts the expression as in the system, the rule is decisively refuted.

Whether we allow ourselves the use of this second procedure is a purely empirical problem. If, for example, on one occasion an informant rejected an expression on the leading-question method and later used it spontaneously, we would doubt the validity of its use in this particular case. In what follows it will be assumed that a leading-question method can be legitimately employed.

In the remainder of this chapter, certain analytic notions of particular relevance to the grammatical analysis of natural languages will be developed. First to be considered are the related concepts of substitution and class. Given an expression of a particular SS and a specified unit within the expression, we can usually obtain another expression in the same system by replacing this unit with another of the same

9. See Rudolf Carnap, *Logical Foundations of Probability* (Chicago, 1950).

kind. The set of units resulting from such substitution, including the original unit itself, will be called a "contextual class." If the unit selected cannot be replaced by any other unit, then it is the only member of the contextual class. This can be illustrated by the following example. Let there be a total of ten permitted expressions of element length 3 in a given system, L_1, as follows:

1. *a p x*	6. *b q y*
2. *a r x*	7. *b r x*
3. *a r y*	8. *c p x*
4. *b p y*	9. *c p y*
5. *b q x*	10. *c s y*

If we consider *x* in the expression *apx* (1), we see that it cannot be replaced by any other element to obtain another expression of the system. Thus *x* by itself forms a contextual class with a single member. A class of two members arises if we hold *a–x* constant in the same expression *apx* (1), since *r* may replace *p*, producing *arx* (2). Therefore, this context class has *p* and *r* as its members. If, instead, we start from *bpy* (4) and hold *b–y* constant, we obtain a class containing *p* and *q* because of *bqy* (6), but excluding *r*. It is therefore seen that, in the same position, different classes arise, depending on the particular expression taken as the point of departure. A contextual class needs for its determination both the particular unit to be replaced by others and a context, namely, the expression in which it occurs. This is the reason for the choice of the term "contextual class." Other types of substitution leading to other kinds of classes will be considered later.

Another key concept is that of a "construction." If we describe a set of permitted expressions of the same unit length within an SS by specifying the class of units which may appear in each position, we are employing the basic notion of construction as used in grammar. Expressions belong to the same construction if they have members of the same classes in the same positions. Thus in English, with words as units, *John sees the house* and *William catches a ball* are members of the same construction. Since we are often concerned with relations among expressions of different unit length which are related in certain specified ways to be considered later and which have a similar internal structure, it will be expedient to extend the term "construction" to such cases. Thus *John sees the house* and *John sees the large house* will be members of the same construction although of unit length 4 and 5, respectively, while *Mary's dress was green*, although of unit length 4, belongs to a different construction. The term "subconstruction" will be reserved for expressions of the same unit length which belong to the same construction. Systems which, like natural languages, have no expression of maximal length cannot be described in terms of subconstructions alone, since we would have to define the subconstructions of each length, leading to an infinity of definitions.

The set of expressions of a given unit length of a particular SS can always be described in terms of subconstructions, though this will not always be the most expedient procedure. It should also be noted at this point that the description by subconstructions is relative to the procedure used to determine classes, the method

of contextual classes already described being but one of a number of alternatives. The following distinctions which are applicable to subconstructions of any kind will be employed. If the class in each position of the subconstruction has the same membership, the subconstruction will be termed "homogeneous," otherwise "heterogeneous." Linguistic subconstructions are always heterogeneous. Subconstructions in which all possible sequences involving members of the successive classes are expressions of the system will be called "perfect," otherwise "imperfect." Thus, if a subconstruction of element length 2 is described as consisting of the class $\{a, b\}$ followed by the class $\{c, d\}$ and if *ac*, *bd*, *ad*, and *bd* all occur, the subconstruction is perfect; if any one of these is not an expression, then it is imperfect. Description in terms of imperfect subconstructions is not very useful because all non-occurring sequences must somehow be specified in addition to the rule of the subconstruction. Finally, it is important to distinguish those analyses in which two expressions of the same length cannot belong to two different subconstructions, that is, in which all the subconstructions are mutually exclusive from those in which this condition does not hold. Those in which the subconstructions are mutually exclusive will be called "unambiguous"; those for which this is not true, "ambiguous."

The subconstructions arrived at by the use of contextual classes as defined above are, in general, imperfect and ambiguous. If, for example, in L_1 we start with *apx* (1) we generate the initial contextual class $\{a, c\}$ because of *cpx* (8). In the second position, the class is $\{p, r\}$ because of *arx* (2). In the final position we have the class consisting of x only. The subconstruction consists, then, of the classes $\{a, c\}$, $\{p, r\}$, and $\{x\}$ in that order. It is an imperfect subconstruction because *crx* does not occur, and it is ambiguous because if, for example, we started with *arx* (2) as our initial expression, we would derive a subconstruction consisting of the succession $\{a, b\}$, $\{r\}$, and $\{x, y\}$ and the expression *arx* (2) belongs to both this subconstruction and the one described above which employed *apx* (1) as its starting point.

Methods of defining other than contextual classes will now be considered. An operation to be called "horizontal transformation" is introduced at this point. Any expression Y will be said to be derived from another expression X by a horizontal transformation, in symbols, $X \rightarrow Y$, if Y results from X by replacing a single unit in X by another one of the same kind. Thus, in L_1, *bqx* (5) $\rightarrow$ *bqy* (6). This relation is, of course, symmetrical: if $X \rightarrow Y$, then always $Y \rightarrow X$. If we apply a succession of transformations, beginning, as before when contextual classes were formed, with *apx* (1) of L_1, we generate *arx* (2) by *apx* (1) $\rightarrow$ *arx* (2) as before, giving us p and r in the second position, but the chains *apx* (1) $\rightarrow$ *cpx* (8) $\rightarrow$ *cpy* (9) $\rightarrow$ *csy* (10) produce q and s also in the second position, so that the entire class consists of $\{p, q, r, s\}$. The same class will eventuate by the method of horizontal transformation, regardless of which expression in L_1 is taken as the starting point. Such a class will be called an "extended class." Similarly, the extended class of the first position will have the membership $\{a, b, c\}$ and that of the third position $\{x, y\}$. All the expressions of L_1 will belong to the subconstruction consisting of the succession of these three classes $\{a, b, c\}$, $\{p, q, r, s\}$, and $\{x, y\}$. That is, by horizontal transforma-