MATHEMATICAL ANALYSIS OF LANGUAGE

ZELLIG HARRIS

Logic, Methodology and Philosophy of Science VI. Proceedings of the Sixth International Congress of Logic, Methodology and Philosophy of Science, Hannover 1979, pp. 623–637

MATHEMATICAL ANALYSIS OF LANGUAGE

ZELLIG HARRIS

It is possible to develop a theory of language in which mathematics plays a role different from its use in most of natural science. This difference is not because language is a human and purposive product rather than a part of objectively given nature. Indeed, language does not have to be studied initially as a human endeavor, with what tools are available for such studies. Instead, one can first look upon language as an aggregate of occurrences of speaking and writing. There are reasons for doing so. Since the intent of the speaker and the effect of speaking can be studied only imprecisely, the possibilities for precise analysis lie rather in the physical events of speaking and writing, as combinations of sounds and of letters, taken in the first instance as events which are characterized not by their relation to the human world but purely by their internal structure.

The problem then is to characterize the events of speaking and writing, that is, to state the parts and their combinations, or the elements and their relations, or other properties that hold for all such events and only for them. Some considerations seem clear from the start. The fact of alphabetic writing shows that discrete objects, the letters of the alphabet, suffice for a representation of language, excluding unspecified expressive inflections. And the experimental method called 'phonemic analysis' shows that the continuous flow of sound in each word can be represented as a succession of discrete phonemes, with just a small set of phonemes sufficing to distinguish the many sounds which occur in speaking a language.

The first picture that one obtains of the structure of these events is that they can each be segmented into successive sentences, with each sentence representable as a sequence of words, or of stems with affixes (all of these called 'morphemes'), and each word or morpheme representable

as a sequence of phonemes. This was the framework of structural linguistics. The next step was to state regularities in the word-successions that constituted sentences (as against those that did not), recognizing for example that *the man walked, the man came, the man left* are sentences while **the man hotel, *the man universe* are not. Each of these regularities holds over a particular domain of words, and the domains of many regularities are coextensive or almost so. To take a simple case, the words that appear before *is here, is missing*, etc., to make a sentence include *the pen, the light, the fork, the knife* (but not *the knive*), while the words that appear before *-s are here, -s are missing*, etc., include *the pen, the light, the fork, the knive*, but not *the knife*. The two domains can be made identical, if we consider *knive* to be a variant form of *knife*, the occurrence of the variant being determined by this *-s*; then the same list occurs before *is here* and *-s are here*. In another type of case, we have *work, think, sing* and many other words all appearing after *I, you, we, the children* and many other words, to make a sentence, but *am* only after *I* (in *I am*) and *are* only after the other words (in *You are, We are, The children are*). Here again, the domains can be made identical if we take *am* as the post-*I* variant of *are*, so that the pair *am, are* form a single member of the domain which includes *work, think, sing*, and which forms a sentence after *I, you*, etc. Such example are best stated if we deal, for the time being, only with very short sentences.

When such regularizations of domain are carried out to the fullest extent permitted by the actually occurring combinations of words, we reach a structural description in which a great many of the exceptions so characteristic of grammar are eliminated. More exactly, these exceptions are transferred from creating restricted domains to creating variant-pairs within regularized domains.

Given this structural picture, another method, called 'transformational analysis', provides a further simplification in the formulation of how word-successions make sentences. This analysis arises as follows: In the structural description referred to above, we can say, for example, that a sentence results if a word of one set *A* (a domain of many regularities) is followed by a word of another set *B*. Thus, a word from *I, children, motors*, etc., followed by a word from *work, sing, think*, etc., makes a sentence. However, not all combinations are equally likely to occur: *I work, I think*, etc., are reasonably likely to be said, as is *Motors work*, and less so *Motors sing*, but hardly *Motors think*. Grammars never tried to specify the individual word-combinations that are reasonably likely to

occur in sentences, as against those combinations that are not, because the data was far too complex and shifting, especially for long sentences. But the problem can be reduced, yielding by the way a new picture of the structure of sentences. This is done as follows:

Consider first the question of long sentences. If we look at the word-combinations in a long sentence, we often find that the sentence can be segmented into parts each of which contains the word-combinations of short sentences. For example, in short sentences, *the book* occurs before *fell, was lost, cost* $ 5, *interested me,* (or after *I found, He bought,* etc.) but hardly ever before *slept, drank wine, coughed.* In longer sentences, containing *the book which,* the words immediately after *which* are from the first set and not the second: *The book which was lost cost* $ 5, *The book which cost* $ 5 *was lost, I found the book which was lost, The book which I found interested me,* but not **The book which coughed cost* $ 5. Indeed, we find in these sentences two occurrences of the words that can occur with *book.* We then say that these sentences are each formed out of two shorter ones each containing *book*; *The book cost* $ 5; *the book was lost* with a change of the repeated noun (*the book*) to *which,* and moving of the second sentence to after the first occurrence of the repeated noun (the antecedent of *which*). This provides a new analysis of *The book which was lost cost* $ 5, not as a sentence containing a relative clause, but as a sequence of two short sentences of which the second underwent certain changes. The concept of 'relative clause' can thus be eliminated from grammar.

The similarities of word-combination can be found not only in different segments of a single long sentence, but also as between two structurally different short sentences. Thus if we consider verbs that appear both as transitives (with a noun object: *He reads poetry, He sells books*) and intransitive (with no object: *He reads, This book sells*), we find that for some verbs the intransitive cases always have the same subject as the transitive, and for other verbs the subject is always one of the objects of the transitive: thus one hardly says *The oyster reads* as one hardly says *The oyster reads poetry,* and there is hardly *The universe sells* as there is hardly *He sells the universe.* We can say that the intransitive cases of these verbs are not independent sentences, but are simply the transitive sentences plus a change: either zeroing the indefinite object, so that *He reads things* becomes *He reads,* or else zeroing the indefinite subject and replacing it by the object, so that *One sells such books easily* becomes *Such books sell easily.* In such ways many distinct sentence structures are analyzed

as being simpler known structures plus stated changes. The changes were called (partial) 'transformations', because they were mappings within the set of sentences, from simple sentences to changed ones, and from pairs of sentences to single long ones.

That these transformations were not merely a simpler way of describing the structure of sentences, but also a real property of them, is seen in the fact that word-sequences which have two distinct meanings, not due to different meanings of their words, can be explained as degeneracies in the transformations; different changes on different base sentences. Thus *Robert Frost reads smoothly* is obtainable, in one meaning, from *Robert Frost reads things smoothly*, and in the other meaning from *One reads Robert Frost smoothly*.

These transformations can be discovered for each language, as being the differences in form between two sets of sentences, roughly when the inequalities of likelihood of word-combination in one set are preserved in the other (as when we compare the likelihoods of words after *book which*, above, with the likelihoods of words after *book*). When we consider the whole set of transformations, we find as will be seen below, that they can suffice to derive the sentences of the language from a subset of short sentences. However, the variety and number of the transformations is too great for them to be fundamental elements of language structure, and indeed it has proved possible to define a very few elementary changes in sentence form, each taking place in a priori statable conditions, such that every transformation is an ordering of one or more of these changes.

These elementary changes are of few physical types: mainly, reduction of a word to zero (e.g. *I* in *I turned and left* from *I turned and I left*), reduction of a word to a pronoun (e.g. the second *the book* to *which*, above), reduction of a word to an affix (to take a very simple case: *-hood* in *childhood* from an earlier free word *had* 'situation'). They are defined as taking place in the word *A* last entering a sentence (in the sense given below) or in the words *B* entering last before that, if the amount of information that *A* brings to the sentence over and above *B* is exceptionally small. Since the set of such *A*, *B* word-pairs is finite, though large, the individual reductions, which are a subset of this set, can be listed for a given language. In contrast, the set of transformations, taken as differences between subsets of sentences is not statable a priori, and possibly is not formulatable in a finitary manner. Note that the set of sentences is unbounded since there is no longest sentence. Very many of the elementary changes are optional; that is, the unchanged 'source' sentence in sayable as well as the changed

one: *I found a book*; *I had lost the book*, as well as *I found a book which I had lost*. We can make virtually all the remaining changes optional if we accept their 'source' sentences as grammatically possible (marked †) though not actually said, e.g. if we take *His early childhood was unhappy* from *His early situation of being a child was unhappy*, reduced to the compound-noun form †*His early child-situation was unhappy*, reduced to a suffix form *His early childhood was unhappy*.

The grammatically-possible sentences satisfy the rules of syntactic structure of all the other sentences, but are not said, either because of special difficulties with particular words (e.g. the words' having dropped out of use in free position, as with *had*, above, or their inability to carry particular suffixes), or because of stylistic preferences for the reduced forms.

Some of the unreduced sentences, i.e. those which are not the product of any reduction, are reconstructed from reduced sentences, as being their grammatically-possible but unsaid sources. Most are simply the sentences of the language before the optional reductions have taken place. Together, the unreduced sentences form a base for the set of sentences of the language, since all the other sentences of the language are formed from the unreduced ones by the regular application of the stated reductions. The sentences that are actually said, both unreduced and reduced, do not by themselves form a well-defined set: many are marginal (e.g. *The baby gave a crawl*), some are dubious (e.g. *I like that she should be on time*), others are said by one person but not by another. But when we include the reconstructed ones, then the unreduced (base) sentences, both those that are said and those that are reconstructed, form a well-defined set consisting of all the sentences that satisfy a certain structure, stated below. The reductions, which create out of these all the remaining sentences, take place over stated domains of the words entering a sentence. Some of these domains can be extended by the speakers, or have other imprecisions, and it is this that makes the remaining, reduction-bearing, sentences a not well-defined set.

This base set of sentences has many important properties. When we include the reconstructed sentences (marked †) in the base, then for each reduced sentence in the language there exists a base sentence which contains no reductions. Since the reductions can be seen to make no change in the information of the sentence, as in the examples above, the information they carry is carried also by their source sentence, so that all the information carried by the language is carried by the base set of sentences.

As we analyze a longer sentence by discovering what reductions they contain where, we in many cases decompose the longer sentence into shorter ones. Thus, recognizing that the *which* above is a reduction of *book* involves admitting two short sentences as the source of *The book which was lost cost* $ 5. If we now consider the longer sentences that remain in the base, e.g. *That John writes music is probable, Mary said that John writes music*, we find that they contain a shorter sentence as proper part (*John writes music*) together with residues which are not themselves whole sentences (*is probable, Mary said*). One can see, however, that the relation of the residual words to each other or to the contained sentence is much the same as the relation among the words of the contained sentence. We therefore try to formulate that relation.

For stated sets X, Y of words, we define a relation $X > Y$ among the words of each sentence, which holds if the necessary (but not sufficient) condition for the presence of X in a sentence is the presence in it of some word of the set Y of which Y is a member. We say that X depends upon Y, or that X is later than Y in entering into the composition of the sentence. For example, certain words A, e.g. *probable, possible, continue*, occur only in sentences in which there occurs a word from a certain set B which includes e.g. *fall, write*, and also includes the words of A itself. Thus we have *That John writes music is probable, John's writing music continues, That John's writing music continues is probable*, but not **John is probable*. This dependence of A-words on B-words may be hard to determine in long sentences, where there may be many words of A and of B present; but it is obvious in the short sentences of the base set, and can then be recognized in all other sentences because the other sentences are decomposable into segments containing the same word-relations as do the short sentences. In contrast to the A-words, words which are in B but not in A can occur also in sentences which do not contain words of A: *John writes music, John fell.* Thus, over the whole language, $A > B$, but $B \not> A$. Another example is words A' such as *entail, because*, whose dependence is on a pair of B-words (rather than a single B); *John's writing music entails his leaving college, John's falling was because of his rushing about, That John's writing music continues entails his leaving college, That John left college is because of his writing music entailing his getting a job*, but not **John entails a job, *Music is because of college*. Here we have $A' > B, B$ (where B includes also the words of A'), but $B \not> A'$. In the base sentences, where these dependences are demonstrable, the dependent word

comes after the first of the ordered words upon whose presence it is dependent: *probable* after the sentence containing *writes*, *entails* between that sentence and the sentence containing *leave*. (Above, the *-ing* and *that* are markers indicating that there is present some word—such as *probable*, *entail*—which is dependent on the *-ing*-bearing or *that*-bearing word.)

The situation of this dependence is less clear in the minimal base sentences, i.e. those that do not contain any shorter base sentence as a proper part: e.g. *John writes music*. Here the dependence is mutual: no word seems to be more dependent than the other. Nevertheless, there is a difference among them, for the second word is similar in morphology and position to the dependent words above. Hence it is convenient to consider the second words of the minimal base sentences, such as *writes*, *falls*, *leaves*, *loses*, *finds*, to be the ones that are dependent upon the presence of their neighbors in the minimal sentences, such as *John*, *music*, *job*, *college*, *book*.

We now consider the structure of the base sentences with respect to this dependence. If we characterize words only by a partially-ordered dependence (and not a mutual dependence), then there must exist some words whose presence in a sentence does not depend on anything, for otherwise no other words—those that depend upon the presence of something else—could be present. By the same token, every base sentence must contain at least one such word. We call these words 'primitive arguments', N: *John*, *music*, *book*, etc. (But not all nouns in a language are such.) Then there must be some words whose dependence is only on primitive arguments for no other kind of word could enter a sentence which contains only primitive arguments. In English we find certain words whose presence depends upon the presence of one primitive argument, e.g. *fall*, *sleep*, *cough*, as in *John sleeps*, etc. Using O_Z to indicate words that depend on Z, we indicate these last by O_n. Other words (O_{nn}) depend upon two ordered primitive arguments, e.g. *write*, *lose*, *find*, *get* in *John writes music*, etc. A few (O_{nnn}, etc.) depend upon three or more—e.g. *put* in *John puts the book on the table*. The words which depend on something are called 'operators', and the words on which they depend, in a given sentence, are called their 'arguments' in that sentence. The symbol for an operator carries subscripts indicating its arguments. The words whose arguments are only primitive ones—the operators considered above—are called 'elementary operators'. For every m-argument elementary

operator in a base sentence there must be present m primitive arguments which are free for that operator, i.e. which have not been counted as arguments of any other operator in the sentence.

In addition there are certain words (non-elementary operators) whose presence in a sentence depends upon the presence of one or more operators. These include the O_0-words, such as *probable, continue* in *That the book fell is probable, John's writing music continued, John's sleeping continued*; the argument of the O_0-word is an operator, which has its own arguments with it. The non-elementary operators include also the O_{00}-words such as *entail, because*, which have two operators as their arguments. English also has words which depend on a pair of arguments, one an operator and the other a primitive argument, in one order or another: O_{n0}-words such as *know, hope, say* in *John knows that the book fell*, etc., and O_{0n}-words such as *astonish* in *The book's falling astonished John.* When an operator becomes an argument of a further operator, it receives *that* or *-ing* as indicator of its changed status. A fact which is of great importance to language structure is that the words which are dependent on operators make no distinction as to what is the argument-class of the operator which has become their argument. For example, *continue* does not ask whether its argument—*write, sleep, continue*, etc.—is an O_n or an O_0 or an O_{n0}, etc.; i.e. it does not depend on the argument of its argument. Thus the condition for the presence of *continue* can be satisfied not only by *write* but also by any other operator of whatever kind: something's continuing can continue, something's entailing something can continue, someone's knowing something can continue. This is one of the facts that make all properties of grammar involve no more than the relations between an operator and its arguments. And, as will be seen below, it contributes much to the mathematical character of the structure.

The base sentences can be formulated in such a way that they involve few or no word-subsets, other than N and O. If we ask what sentences are possible rather than which ones are actually said, we not only admit such cumbersome sentences as *His early situation of being a child* but also the sentences with unlikely combinations of words within the normal grammatical constructions, e.g. *He took a crawl, The astute ceiling thinks that we are late.* In the grammatical statement there is no point at which one could draw a line between the set of likely combinations (e.g. *He took a walk*) and the set of unlikely ones, nor are the likelihoods unchanging. For the base sentences, if we say that a certain set of words are, say, O_{n0}, i.e. that each requires the presence of a primitive argument N and an

operator *O* as its ordered arguments, then each of them can occur with any *N* and any *O*, as in the example of *Motors think* above; and indeed that sentence could occur in science fiction or in a joke, without being ungrammatical. Within each set, e.g. O_{no}, each operator has a partial ordering for its likelihood of occurring with each word in its (*N* or *O*) argument domains. This likelihood is imprecise, but it is preserved under all further events in the composition of the sentence, whether reductions or the entry of further operators. Indeed, if two sentence-forms show the same inequalities of likelihood for the arguments of an operator, we assume that one is the result of reductions (transformations) in the other. The restrictions and exceptions that are so familiar in grammar can be stated as limitations not on word-entry but on the domain of reductions: which words get reduced under what conditions.

The whole grammar is thus stated in terms of the operator-argument (or dependence) relation. Every occurrence of an operator on the sentence thus produced produces a further sentence in which the first one is an argument, and a proper part. As to reductions, in almost all cases, the word that gets reduced is one which contributes to the sentence little or no information, given its position over its arguments or under its operator in the sentence as so far constructed. The reductions are made upon entry; that is, (a) on a word as it enters the sentence, or (b) on one of the last entering words as an operator enters upon them, or (c) as soon as a further operator empties the informational contribution of the given word. As an example of (a): in *I request you*: *wash yourself!* the operator *request* with its first two arguments *I, you* can be zeroed, leaving the third argument *Wash yourself!*; in this rare kind of reduction, the informational grounds are the performative status of *I request you*, namely that saying *I request you* of an imperative is the same as making (saying) that imperative sentence. As an example of (b): the indefinite nouns (or so called 'pronouns') *something*, *things*, etc., carry little information; hence, in most cases, when they are the second argument of an operator they are zeroable, as in reducing *John reads things* to *John reads*. As an example of (c): *Boys take these jobs because of needing the money* is said only if it is the same boys that need the money; it is therefore reduced not from *Boys take these jobs because of boys' needing the money* but from something like *Boys take these jobs because of boys' needing the money*, where the first argument of the second argument is the same as the first of the first (i.e. where the second-mentioned boys are the same as the first). Thus it is only after the metalinguistic last sentence is added that the second

boys is zeroable. That reductions are made as soon as the conditions for them are satisfied, and are not otherwise delayed, is seen in many sentence derivations, and explains for example why pronouns are late changes in a sentence, since they depend upon that sentence being joined to another sentence and are therefore formed after the internal changes of each component sentence have been made.

In respect to meaning, the operators are predications. That is, the word whose presence depends upon certain other words says something about those other words. Those words and affixes of English which are not obviously operators or their arguments—e.g. *the*—turn out to be derivable by reduction from particular operators and arguments. All relations other than the predicational operator-argument relation—e.g. the modifier relation—can be obtained via particular reductions from the operator-argument relation. The meaning of a sentence, or rather the information carried by it, is given directly by the meaning of each of its operator-argument portions, i.e. by its elementary operators as predications on their arguments, and then by each successive further operator. Then the meaning of a sentence is not something else again, to be considered after the syntax is determined, but correlates in a regular way with the syntax of the sentence.

And now, as to the mathematical possibilities. It is possible to apply mathematics, in the usual ways, to the study of language phenomena. One can describe stochastic processes for determining word and sentence boundaries, and certain algebraic structures for sentence composition. As generally in applied mathematics, these investigations accept certain objects which are determined within a science of the real world—linguistics; they then describe the combinations or changes of these objects. However, the analysis given above makes possible something else, a mathematical characterization of language. The way to this is prepared by the elimination of restrictions and exceptions from the occurrence-dependence of words, moving these to the domains of reductions on the words. This makes it possible to consider mappings between sets of linguistic objects—at least in the unreduced sentences—without having to formulate special provisions for exceptions and the like. A further step here is the fact that the only arguments which characterize the various sets of operators are N, the primitive arguments, and O, the set of all operators. The importance of this is that these arguments are themselves defined purely by their occurrence-dependence. N is the set of words whose presence in a sentence is defined as not depending on anything. Within the vocabulary of the

unreduced sentences, O is the complement set, of words whose presence depends on the presence of something else—N or O in some combination. Thus, words are characterized in sets either as having null dependence (N), or as depending on one out of a few combinations of words which are in turn identified only as having null or non-null dependence. Since the word-sets are not otherwise defined, they are characterized only by their relation to words which are characterized in respect to this same relation. Within each word-set, the individual words can be identified syntactically by their inequalities of likelihoods of occurrence in respect to the individual words in their operator-set, whose words in turn can be identified by their inequalities of likelihoods of occurrence in respect to the individual words in their argument sets. We are thus dealing with sets of arbitrary objects, defined only by their participation in a relation in respect to each other—a mathematical object.

Language is a particular realization of this mathematical object with its occurrence-dependence relation, a particular interpretation of the abstract system. But any other physical system in which the combination of parts was based solely on such an occurrence-dependence relation would be language-like. And if the occurrence-relations of the new physical objects are identical in detail with those of language, we obtain a set of sequences isomorphic to the set of sentences, as indeed we have in writing vis-a-vis speech.

The structure of sentences and the relations among them can be described as certain simple algebraic structures. These are chiefly partial orderings, monoids of non-elementary operators (but not of the binary O_{00}, which are mostly non-associative in respect to meaning), and equivalence relations which provide partitions of the set of sentences. These structures are important, because every relation in them has an interpretation which is an essential part of the meanings of sentence structures; and those meanings in a sentence which are directly connected with the grammar of the sentence are interpretations of the stated relations in these algebraic structures.

The sentences of the base set are a partial order (a particular kind of semi-lattice) of arbitrary objects. The objects (in actual languages, words) have the additional property of being classifiable according to whether they occur, in the operator-argument semi-lattices (i.e. in sentences), as (1) l.u.b. only of their own occurrence in the partial order (N, elementary arguments), or (2) l.u.b. only of the latter and themselves ($O_{n\ldots n}$, elementary operators), or (3) l.u.b. also of objects which are themselves the l.u.b. of

objects other than themselves, as well as possibly of N (these are $O_{\ldots O \ldots}$, non-elementary operators). That is to say, these three types of l.u.b. positions are filled in general by different objects. Furthermore, within each of the latter two position-types there are sub-types according to the number and order of N and O in the immediately lower position in the partial order: in type 2, according to how many N are immediately below it; in type 3, according to what sequence of N and O is immediately below it (O representing any object in type 2 or 3). These sub-types of l.u.b.-status are generally filled by different objects.

The set of all unary non-elementary operators, i.e. those whose argument-dependence includes precisely one operator, generates a free monoid, with successive application (i.e. next later entry) as operation, and the null operator as identity. In this, the monoid-words are products of operators $O_1 O_2 \ldots O_n$ (where $O_i O_{i+1}$ means that O_{i+1} is the operator on O_i in a sentential partial order). A product of two monoid-words is itself a monoid-word; the multiplication is associative. Each monoid-word represents the succession of operators on an elementary sentence or on a binary operator on two sentences. This structure has not so far been found to be of any great importance in dealing with language. However, it illustrates how using the partially-ordered dependence-relation, instead of the overt word-sequence, makes it possible to find various mathematical structures in language. In contrast, word-concatenation in sentences is non-associative and ambiguous: *The yellow and green cards* can be derived both from *The cards which are yellow and green* and *The cards which are yellow and the cards which are green.* The entry of operators, which together with reductions describes the same sentences as concatenation would, is associative and non-ambiguous. Mappings and operations on sets of sentences can therefore be more conveniently carried out on the entries, and in particular on the operators, in the sentences than on the word-sequence of the sentences.

The binary non-elementary operators, i.e. those whose arguments include two operators, form a set of binary compositions on the set of sentences. In the base set, each binary non-elementary operator can act on every pair of sentences, although its likelihood of occurrence is lower on sentence-pairs which do not contain in their base form a word in common. In the whole set of sentences, certain statable pairs (e.g. an assertion and a question) will not appear under certain binary operators: **I am late because will you go?* In these cases we can say that the product of the two sentences (*I am late*, *Will you go?*) under the binary operator (*because*) is included

in the null sentence. Products of these binaries are in general not associative.

The reductions in a sentence act as a partially ordered set on particular operator-argument pairs, those which have the likelihood (or low information) properties required for the given reduction. Some of these reductions (if there are more than one) can be viewed as taking place simultaneously on the given operator-argument pair; others are such that one reduction operates on the resultant of another on the same operator-argument pair. Some of the large grammatical transformations such as the interrogative form and the passive are not single reductions but successions of reductions on a single operator-argument pair in a sentence or on successive operators in a sentence.

The most important algebraic structures in the set of sentences S are those which arise from equivalence relations in S in respect to the particular operator-argument semi-lattice in each sentence, and in respect to the highest operator (the upper bound of all words in the semi-lattice) or to the reductions on words in the semi-lattice. These equivalence relations identify the informational sublanguage (the base) and the grammatical transformations, as will be seen below.

We note first that the resultant of every operator is a sentence. Every unary non-elementary operator acts on a sentence (and possibly some N) to make a further sentence; and every binary non-elementary operator acts on two sentences (and possibly some N) to make a sentence. Every reduction acts on a sentence to make a (changed) sentence. All of these, in acting on a sentence, preserve the inequalities of operator-argument likelihoods in the operand sentences. The unary non-elementary operators are a set of transformations on the set of sentences: each maps the whole set of sentences S into itself (specifically, onto a subset of sentences which have that non-elementary operator as their latest entry); and the binaries map $S \times S$ into S. The reductions are a set of partial transformations on S, each mapping a subset of S (sentences containing a particular low-information operator-argument pair) onto another subset of S (sentences containing the reduction on a member of that pair).

The preservation of inequalities of likelihood under transformations, i.e. under the non-elementary operators and under the reductions, is of great importance. Without it, there would be no semantic connection between a sentence and its occurrence under further operators or reductions. The operators preserve the likelihood-inequalities and the meanings in their operand sentences, although with a reasonable number

of specified exceptions. The non-elementary operators also add their own meanings and likelihoods in respect to their argument, so that the inequalities among the resultant sentences (with their new higher operator) need not be the same as among the corresponding operand sentences. As to the reductions, they preserve with only few if any exceptions the inequalities of likelihood and the meaning of their operand sentences and add no objective information to it; they are paraphrastic. Here too the reduction may raise (or lower) the likelihood of occurrence in the resultant sentences, but for the most part equally on all its operand sentences.

We now consider the set S, where each word-sequence which is grammatically ambiguous in n different ways is considered to be a case of n different sentences. S is a semi-group under the binary operator *and*: for any two sentences A, B we have A *and* B as a new sentence C. (This, after adjustments are made for the stable A, B pairs which do not take *and*.)

We present now a structure which isolates the minimal subset of the set of sentences as a residue of the non-elementary operators and reductional transformations. It has little importance when the great bulk of transformations are products of such simple reductions as have been established for English, and when the minimal sentences can be characterized, as has been done here, as the resultants of elementary operators on primitive arguments. However, in a language in which we do not have so clear a picture of the structure of the set of transformations or of the set of base sentences, such as a way of identifying the minimal sentences is useful. To obtain this structure, we take an equivalence relation in S, whereby two sentences are in the same equivalence class if they contain the traces of (i.e. exhibit the presence of) the same monoid-word of unary operators and the same partial orderings of particular reductions. There is a corresponding binary composition in the set of equivalence classes E, with E_A *and* $E_B = E_{A \text{ and } B}$ (where E_X is the equivalence class to which the sentence X belongs). In the natural mapping of S onto its quotient set E, the kernel of the mapping, i.e. the sentences which are mapped onto the identity of E, includes elementary sentences and also the resultants of the binaries. In each of these resultants, the two operand sentences are then assigned to equivalence classes in the same manner as the original sentences. When no resultants of binaries are left in the kernel of the natural mapping, this sub-kernel contains only minimal sentences.

A different and much more important structure is obtained if in the set S we take an equivalence relation by which two sentences are in the

same equivalence class if they have the same ordered word-entries (i.e. the same operator-argument semi-lattice). Since almost all reductions are optional, each equivalence class contains (with possibly certain adjustments) precisely one reduction-less sentence. The set of these is the base set, from which the other sentences are derived by reductions. The base set is closed under the word-entry operation: any word sequence satisfying this form is such a sentence. Hence we may call this set a 'sublanguage'. Since the domains of successive reductions are monotonically decreasing, the base set of sentences, one from each equivalence class above, is the most unrestricted in respect to word domain.

There are many properties of language that can be derived from this analysis, or are clarified by it. One is that since the base set, which suffices for all the information carried by language, has virtually no restrictions or exceptions and lacks all the special constructions of grammar such as conjunctions, tenses, etc., it follows that, contrary to common views, all these are not essential for expressing the information carried in language, nor are they essential for language. Another is that since the whole structure of language is seen to be predicational, it is clear that language developed as a tool for communicating information rather than purely as a form of expression. Yet another is that since the whole of language arises in explicable ways from so simple a relation as the dependence of word-occurrences, there is no need to assume any inexplicable structuralism underlying language.