

ist, ist die Bedingung von "useable communication" in Subsprachen präziser erfüllbar als in der ganzen Sprache. Die Grammatiken der Subsprachen bilden einen Durchschnitt mit der Grammatik der ganzen Sprache, und die Vereinigung der Sätze aller Subsprachen ergibt die Sprache. Die Beziehung zwischen den verschiedenen Satzmengen und ihren Grammatiken ist noch weiter zu untersuchen.

[*Senta Plötz*]

NEW VIEWS OF LANGUAGE

Zellig Harris

1. RESULTANT OF OPERATOR SYSTEMS REPLACES SINGLE DESCRIPTIVE SYNTAX

We start with established analytic procedures of linguistics, namely, separating out the complementary and free-variant relations among elements (in respect to their environments) from their independent combinational limitations (i.e. from the other — residual — restrictions upon their environments); but we carry these procedures as far as possible, given the particular vocabulary of the language.¹

This work leads to isolating certain elements, or disjunctions of elements, which are relatively unrestricted as to environment (although they have the residual restrictions noted above). Many of the elements have free variants over proper parts of their environments; that is, these free variants are restricted to occurring with stated subdomains of environments. (And in the disjunctions of elements, each member is restricted to its characteristic environment.) When this work is carried out on a transformational analysis of language, the elements are transformational operators, and primitive arguments of transformational operators; and the relevant environment of element A is merely the set of operands of the operator A (immediate, or perhaps in some cases also deeper down; we may in some cases also consider the operators on A as environment of A). The residual restrictions are: that certain classes of these "unrestricted" elements occur only in particular positions relative to certain other classes (i.e. have only these other classes as their immediate operands); and that the members of the operator class impose an acceptability ordering on the members of the immediate or deeper operand class.

As we isolate the unrestricted elements, we can also state the relations between each of them and their free or complementary variants (including zero and pronouns). And as we note the distinguished sentence positions of these unrestricted elements (*vis-à-vis* their operands) in their descriptively simplest sentences, we note also the various other positions in which these elements appear in various syntactically-identifiable paraphrases of their descriptively simplest sentences. These other positions, as also the change over to the variants of the unrestricted elements, we then attribute to certain operations which act upon the unrestricted elements in their "original" positions.

We then note that the set of unrestricted elements in their distinguished (original) positions constitute a separate grammatical system, I: the operation (or juxtaposing) of these unrestricted elements (of I) upon other unrestricted elements (of I) form the sentences of I (a subset of the sentences of the language). And the various free and complementary relations and (later) position-changings constitute another grammatical system, M: the operators of M are changes (operations) upon the sentences of I, producing all the remaining sentences of the language. I and M are systems, in that each consists of certain operators and relations among them; each is very simple and functionally interpretable. Together they describe all sentences of the language far more compactly than does any grammar.

These two systems do the same work that grammar does; i.e. they between them describe all sentences. But they are not in any convenient way components out of which the grammar can be said to be composed. That is, when we compare them with the grammar we do not find that the various items of the grammar are simply distributed among these two systems. Indeed the two systems operate in an unexpected manner, in that the prior system I describes certain sentences, some of which are quite cumbersome and unpreferred sentences (though they cannot be said not to exist in the language), and the second system M operates on the sentences of I to produce the remaining (and often more natural) sentences of the language. If these two systems constitute a characterization of sentences which is superior to grammar, it is not because they are immediately natural generators of the grammar — e.g. as describing the simple sentences first and from these the more complex ones — but because they are functional simple systems, each possessing several important and reasonable properties, which together describe the same thing that is described less compactly by grammar.

If one thinks of the many different rolling and slipping motions of a ball on an inclined plane, one sees that the set of all ball-motions is large and varied, but all of them can be described as resultants of two independently operating motion-describing (or: motion-determining) systems: gravity and friction. In a somewhat similar way, there is a set of all sentence-forms,

all of which can be described as resultants of two independent form-describing (or: form-producing) systems: I and M. In physics, a direct listing of all the ball-motions, and a classification of them by partial similarities, is patently impossible. In linguistics, because the data is discrete and linear, consisting of sequences of a manageable number of sound-types or morphemes, it has been possible to make grammars, i.e. efficient descriptions of the sentence-forms by their partial similarities.

When we replace the grammar by the I and M systems, therefore, we are replacing syntax, which is a classificatory procedure whose criterion is environment, by a theory of operators. The ordered operators characterize the sentences. But that we are now no longer directly describing the shape of sentences, as is done in grammar, is seen in the fact that some of the most prominent items of grammar, such as the paradigms, do not appear as such either in I or in M. The I system will contain various operators, such as *(was) in the past, (was) yesterday, (will be) in the future, (will be) tomorrow*. The M system will contain morphophonemic rules that all (verb) operands of *in the past, yesterday*, etc. receive a suffixed *-ed* and that *in the past* is then zeroable. In some languages the shape of the "past-tense" affix will depend upon certain properties of the first operand (subject) of the verb (e.g. feminine). Thus certain striking regularities of partial similarity in sentences are just the resultant of M operations acting on I operators in the construction of the sentence.

Each of the I and M operators, however, is directly syntactic in the sense that it directly places a morpheme in a sentence, or changes its shape or position there; that is, each operator makes a contribution of a fixed type to the morpheme arrangement which is the sentence. Each operator in I contributes a single element (usually, one morpheme) to the sentence, and always in the fixed position which is the defining position for operators (in English: after its first operand). Each operator in M changes the physical (phonemic) shape, or the sentence-position, of a pre-existing element of the sentence; and all in only a few functionally-interpretable types of change.²

The nature of I and M as determiners of sentence-forms raises the question whether there may not be other respects in which the forms or events of language can be described or produced as resultants due to the co-effect of more than one system.

2. LANGUAGE APPROXIMATES THE STRUCTURE OF DISTINGUISHED SUBSETS

The properties of I and M lead to an additional view about language. If I had been simply a particular subset of sentences, then the fact that many of the sentences of I are cumbersome would be of no theoretical interest. How-

ever, I and M are each structurally-characterized sentence-producing systems, and the operators of M necessarily operate on the resultants of the I-operators (and not vice-versa). Hence I is descriptively prior to M, and the source for it. But in many cases the sentences of I are morphemically more complicated or derivative, and presumably historically not prior, to those of M. E.g.,

His arrival yesterday preceded (or: was before) her departure yesterday.
His arrival yesterday precedes (or: is before) her departure tomorrow.

are both in I. But,

He arrived yesterday before she departed.

is in M. It is not natural to think of the sentences of I as source for M.

This situation can be understood if we think of the special status of I, not as temporally anterior (something which exists, and is thereafter operated on by M), but as a limiting condition on the language: something to which language approximates, toward which it (structurally, not temporally) develops. In the M sentences, the available morphemes are used inefficiently: e.g. there are environments where by custom certain words are not used, even though a useable communication could have been obtained by using them. Thus *before* is not used between two sentences if their tenses (relative to the speaker) differ:

✗ He arrived yesterday before she will arrive tomorrow.

(If *He arrived yesterday, before she will arrive tomorrow* is acceptable, it is by virtue of a source, and *this will have been before*.) The I sentences had been syntactically isolated as being composed of that particular vocabulary which is most efficiently used, i.e. where each morpheme is least restricted in its environments. We see that in the language as a whole there is a condition which limits the occurring sequences of words; this condition is related to meaning but not in a precise way, as will be seen below. (So to speak, only certain word combinations make sense and are useable.) However, there are many restrictions on word-combination which are only indirectly related to this limiting condition (or in any case to meaning), but may be more directly related to custom or to the history of vocabulary use, etc. Thus one could try to figure out the meaning of *before* so as to fit the actual restrictions around it (e.g. to having both tenses identical), and then say that *before* is not used between *yesterday* and *tomorrow* because of this meaning specification (i.e. that *before* expresses objective time-ordering within subjective tense). But aside from the circularity of such a claimed causation, we find that *before* does not have this semantic property in *is before* (which

does indeed occur between *yesterday* and *tomorrow*). And in any case *is before* is less restricted: it occurs with all sentence-pairs in which we find *before* and also in other sentence-pairs. Thus sentences (in I) which contain *is before* instead of *before* participate in a more efficient, less redundant, use of vocabulary (less restrictions on word-combination). The sentences in I are limited only (or almost only) by the limiting condition of useable communication. They can thus be looked upon as something toward which the structure of word-combination developmentally approximates.

It should be mentioned in passing that the vague term "useable communication" is more applicable here than 'meaning'. Explanations of sentence-structure (word-combination) in terms of meaning are ad hoc unless we say that the meaning of a sentence is the sum of the meanings (otherwise determined) of its component morphemes and their arrangement. But this palpably does not hold for metaphors (*fly in His words flew thick and fast*) and for many other special cases and innovations. 'Useable communication' or the like is a social (rather than semantic) term, which recognizes the functional condition or requirement upon sentences while allowing any manageable, finite, number of special cases of word-combination, arising out of whatever historical or social or linguistic situation. (The types and sources of these special cases can undoubtedly be studied.)

Once we have the above case of something (a subset of sentences) being analyzed as a distinguished structure (I) toward which language approximates, we can ask whether there are any other subsets in language which have, so to speak, purer or stronger properties than language as a whole has, such that we might say that the structure of language approximates the structure of these subsets.

3. DISJOINT AND INTERSECTING SYSTEMS OF SUBSETS OF SENTENCES

There is a further observation, which can be better understood in light of the two views noted above. This is the observation that if we make a grammatical description of only the sentences in a very restricted subject-matter, e.g. a subsience, we obtain a grammar which is not quite the same as the grammar of the whole language (and indeed intersects the grammar of the language rather than being simply included in it).³ (The sentences in question must, of course, not be selected semantically but on some controllable grounds, e.g. as being all the sentences in a set of scientific papers.) There are even grounds for a conjecture that certain procedures of theory-building investigation (such as determining acceptability-ordering within sentence-sets), which are very difficult to establish for the language as a whole, may be successful within the set of restricted-subject-matter sentences.

This raises the question of how limitation in respect to subject-matter can have a grammatical — and, even more surprisingly, a theoretical — effect. The existence of such an effect can be understood if we think of the regularities of word-combination as satisfying the conditions of useable communication. Vague as this concept is, it is clear that the word-combinations of useable communication in a restricted subject-matter, especially in a science, will be partially different from those elsewhere, and may be more precisely distinguished from the unuseable word-combinations.

If such a subject-matter sublanguage has a more precise grammar, and if theoretical procedures are more precisely carryable-out there, than is the case for the language as a whole, we would want to look upon the greater success in the sublanguage as throwing light upon the more difficult situation in the language as a whole. But it seems odd to say that language is characterized by a structure which is really establishable only in sublanguages which clearly arose later in time than the language itself. Here, however, we can appeal to one of the views discussed above, namely seeing language structure as the satisfaction of conditions, something toward which language (in particular uses, at least) develops and approximates, rather than as something given and descriptively prior which language then realizes. We would then say that the combination-determining communicational useabilities which are only approximately satisfiable in colloquial language or in language as a whole are more sharply satisfiable (for somewhat different communicational useabilities, of course) in subject-matter subsets of sentences.

Finally, we can ask what is the structure of language: not only the structure of sentences, but also the structure of sentence-subsets (what subsets are of interest, what relations among them, etc.) We have already seen that the set of sentences of a language is decomposable into certain disjoint subsets having important properties: I and M. There are also other ways of establishing important disjoint subsets, such as various types of metalinguistic sentences. In addition, we see here the possibility that there are subsets of sentences with possibly non-empty intersects, each containing the sentences used in particular subject-matters, such that each of these subsets has a particular grammar (though all the grammars have many properties in common), and such that the determination of these grammars establishes the purest form of linguistic structure, and such that the language as a whole is the union of all these subsets (and of any residual sentences of the language). The relation of the grammar of the whole language to the grammar of the sublanguages (and of any residual subset of sentences) is, of course, unclear at present.

NOTES

1 One could go yet further in finding free-variant (i.e. freely alternating) relations among morphemes if one considered not morpheme-bounded phoneme-sequences but particular environmental occurrences of these. Two morphemes, A and B, may be free-variants (written \approx) of one another when they occur in environment X (i.e. $AX \approx BX$), but not when they occur in environment Y (i.e. $AY \not\approx BY$). This further step is difficult to control, except in sublanguage grammars (3, below), and is not considered here.

2 The question might be raised: at what point does classificatory syntax get replaced by syntactic (i.e. morpheme-sequencing) operators? When linguistic transformations are understood as a system of partial transformations in the set of sentences, i.e. as a way of deriving one sentence from another, there must always be some residual subset of primitive sentences, the structure of which must be given by a pure (non-transformational) syntax. Transformations then are a mapping from this syntactically-described set onto the whole set of sentences. But when transformations are decomposed into elementary transformations, and these are classified into the I operators occupying second position (for English), and the shape-changing operations of M, we can say that the verb-segment of the elementary sentences is also a second-position operator, hence in I. Then I produces elementary and derived "informational" sentences out of the primitive set of nouns; and M changes these sentences into other forms. In this case, I and M produce all sentences without any prior syntax.

3 In contrast, if a sufficiently small subset A of sentences were selected at random, rather than by subject-matter, the grammar made out of A would in general be a subset of the grammar of the whole language.