TRANSFORMATIONS IN LINGUISTIC STRUCTURE

ZELLIG S. HARRIS

Professor of Linguistics, University of Pennsylvania

(Read April 25, 1964)

EMPIRICALLY, a language is a set of discourses —things said or written. Each discourse can be shown to be a sequence of (one or more) sentences (or certain fragments of sentences), a sentence being the largest stretch of language whose composition can be described in certain compact ways. (It is possible to state additional properties of a discourse, but not—at least at present—direct rules of how it is composed.) There are several ways of analyzing the structure of sentences, and the applicability of one does not falsify the others. The most common method, both in traditional grammar and in modern linguistics, is to describe sentences as composed of certain constituents, e.g., subject and predicate, and these in turn of certain smaller constituents (say, subject as composed of noun and its modifiers; predicate as composed of verb plus object), and so on until we arrive at morphemes (morphologically indivisible words, stems, affixes). This can be stated in a compact hierarchy of rules or mappings, the rules and their hierarchy all showing some regular character. All sentences, or all of a distinguished subset of sentences, are composed in this way. The various constructions, like "noun-modifier," "subject," are only intermediate constructs of the hierarchical operation of the rules.

In contrast, transformational analysis describes sentences as being composed of sentences, rather than of parts which are themselves not sentences. Rather than ask how one can analyze a sentence into parts and those parts further into parts, and so on, we now ask whether there is an interesting way, not *ad hoc*, for decomposing a sentence into sentences, and those into yet other sentences, until one reaches certain elementary sentences which are not further decomposable.

In the case of English, and of the other languages investigated up to now, these elementary sentences turn out to be from two to five words long. To have some picture of this decomposition, consider the sentence: *The book was picked by the man.* It is decomposable, via a particular opera-

tion, into a single sentence: *The man picked the book,* plus passive operator. Similarly, *The man who spoke just left,* would be decomposable into two sentences: *The man just left, The man spoke,* with a *wh*-connective between them. (This is the *wh* of *who, which, when,* etc.)

Now, it is necessary to indicate by what criteria one would determine whether a sentence is or is not composed of another sentence.

Let us take a sentence form, which is to say a sequence of word classes, with possibly certain small grammatical words or affixes such as *is,* which we will call the constants of that form. We take first the grammatical form $N_1\ V\ N_2\ P\ N_3$. (N: noun, V: verb, P: preposition; the subscript numbers are only to identify the various N). There are very many sentences of this form. Some of them are more acceptable, some are less acceptable. Let us consider a number of them— for convenience, all containing the same verb, *mail.*

The man mailed a letter to the office.
The man mailed a letter to the child.
The man mailed a letter to the moon.

The last is questionable, even though it may become a reasonable sentence. However, in this listing we shall not exclude sentences on the basis of meaning. We are seeking here a structural relation among sentences (that is, a relation between sentence structures); and in order to establish this we seek first some usable difference among the sentences of each form. The difference may be in acceptability, in type of meaning, etc. In the sequel we shall see that we do not use the actual fact of whether a sentence is acceptable or not, or what its meaning is, but some relation formulated on this property, a relation which will be preserved in the structural equivalence we are about to define.

We continue, then, with:

The man mailed the moon to the sun.
The office mailed the house to the letter.

and finally:

The idea mailed the moon to cheese, which some people would say is ungrammatical, while others would say it is nonsensical (whatever the precise meaning of "nonsensical" may be).

Among these and any sentences in a form, there are differences in meaning. Meaning, however, is not the best criterion not only because it is difficult to determine, but also because it is too individual: Every sentence has a meaning which perhaps differs in some sense from that of any other sentence.

Each sentence also has some property of acceptability, as being fully natural, nonsensical, barely grammatical, etc. Many sentences are roughly equivalent in this respect, and it is this criterion which we shall consider here. But for our present purposes it does not matter what criterion one uses. Other non-trivial properties, in respect to which the sentences in a form can be scaled, can suffice.

Now, let us consider some other sentence forms containing the same word-classes: for example, $N_1 \, V \, N_3 \, N_2$. We have *The man mailed the office a letter.* This may be a bit less acceptable than *The man mailed a letter to the office,* but this is not relevant, as we shall see.

The man mailed the child a letter.
The man mailed the moon a letter.
The man mailed the sun the moon.
The office mailed the letter the house.
The idea mailed cheese the moon.

All except the most natural sentences of the $N_1 \, V \, N_2 \, P \, N_3$ form may be a bit less comfortable when rearranged into this form. But what is relevant is that the differences in acceptability among the various sentences in the $N_1 \, V \, N_2 \, P \, N_3$ form are preserved in the $N_1 \, V \, N_3 \, N_2$ form.

As I said, it does not matter for our purposes whether one uses acceptability or normalcy of meaning, or any scalable property that one wishes, or whether one uses measures of response, e.g., measuring how long it takes a hearer to recognize the sentence.

We can consider another form, e.g., N_2 *was Ven (by N_1) P N_3,* where parentheses indicate omittability.

The letter was mailed (by the man) to the office.
The letter was mailed (by the man) to the child.
The letter was mailed (by the man) to the moon.
The moon was mailed (by the man) to the sun.

The house was mailed (by the office) to the letter.
The moon was mailed (by the idea) to cheese.

Here, too, any differences among the sentences, as to which are more and less acceptable in this form, are preserved in respect to the $N_1 \, V \, N_2 \, P \, N_3$ form and the $N_1 \, V \, N_3 \, N_2$ form. Recasting a sentence from one of these forms to the others does not affect its difference in acceptability relative to the identically recast sentences in the same forms.

I have proposed the term "transformation" for this relation, since it is a transformation of members of a set into other members of the set, preserving some important properties: It is the rearranging of the words of a sentence from one form into another in which the difference among the sentences of a form, as to acceptability, or as to like properties, is preserved.

Now, let us take another case. Let us take the form $N_1 \, V \, N_3 \, P \, N_2$. Here we have:

The man mailed the office to a letter.

Note that a sentence which was perfectly acceptable in the first form becomes not acceptable, or nonsensical, in this new form. Then we have:

The man mailed a child to the letter.
The man mailed the moon to a letter.
The man mailed the sun to the moon.
The office mailed the letter to the house.
The idea mailed cheese to the moon.

Sentences which differed greatly in their acceptability in the other forms are equally unacceptable in this form; and some sentences which were nonsensical in the other forms are acceptable when rearranged here. That difference between sentences which was found in the first form and preserved in the other two no longer obtains here. Some sentences in this form may be sensible, some may not be sensible; but the differences between them, when their words are rearranged in this way, are not the same as the differences among them in the other forms. This form, then, is not a transform of the others, even though it is the simplest permutation of the first form. It may, of course, be a transform of something else.

Another form which is not a transform of $N_1 \, V \, N_2 \, P \, N_3$ is, for example, $N_1, \, N_2, \, V \, P \, N_3$: *The man, the office, mailed to a letter; The man, a child, mailed to a letter;* etc.

The definition of transformation can be refined to be a relation, not between forms, but among sets of sentences such that each set has a unique

range of forms into which its sentences can be rearranged transformationally. The decision as to whether two forms are transformations of one another becomes far more complicated if we want to consider cases in which some of the sentences (word-choices) in a form are transformed into another form, while others are not. Here there is the difficulty of distinguishing the reasonable transformations from chance sets of sentences whose acceptability-difference is preserved when their words are rearranged into another form. However, internal checks can be made to support the distinction. And the fact that the transformations of a language form a tightly knit structure, as we shall see, shows that they have not been defined in an *ad hoc* manner.

The decision as to whether one sentence is a transform of another is based not upon the individual sentences, but upon the forms which the sentences have and upon the differences in acceptability among the sentences within a form. The decision is not based in a direct way upon the meaning of the words.

Since many people have questioned the attitude of formal linguistics to meaning, I should remark that the avoidance of defining linguistic relations on the basis of meaning is not because meaning is considered to be pointless. It is because we are trying, among other things, to discover a formal basis or correlate to meaning rather than to assume meaning as an undefined linguistic primitive.

It is not always the case, though it often happens, that the sentences in one form mean the same as their transforms in another form. Some people may say that there is a difference in meaning between the passive and the active, hence between *The man mailed the letter to the office* and its transform *The letter was mailed by the man to the office*. But it is very different from the difference in meaning between *The man mailed the letter to the office* and its nontransform *The man mailed the office to the letter*. There are, it is true, transformations which bring in a large difference in meaning. For instance, the question and the negative are transformations, since they simply permute some words of the sentence, add constants, in the same way for all of the sentences of a given form; and this without changing the difference in acceptability. But the difference in meaning which is due to the transformation is the same for all sentences, and does not affect the relative acceptability of the sentences. (Differences of truth result from negation, differently for

different sentences, while the question eliminates any property of truth; but truth is not directly involved in defining transformations.)

So though there may be a change in meaning due to a transformation, the change is either stylistic or subjective, in contrast to the objective difference in information between sentences which are not transforms of each other (as in the example above); or else there is a change of a logical or operational type, which is constant for all the sentences of the form. The difference between (*a*) transformational and (*b*) other changes in meaning, which result when the words of a sentence are rearranged (*a*) transformationally, or (*b*) otherwise, may help in distinguishing two senses of meaning—stylistic and quasi-logical as against substantive—which may be useful for the retrieval and analysis of the information contained in scientific discourse.

Sentential transformations, as we have seen them here, are a relation among sentences. Indeed, they are an equivalence relation, though not exactly on sentences but rather on readings of sentences. (A sentence may have two or more grammatically distinct readings, each of which has transformations different from those of the others. Such a sentence is called grammatically ambiguous.) However, we can define the difference in form between two transforms of each other as a directed operation producing one out of the other. This is useful when we speak in terms of decomposing and composing sentences.

In English, and in the few languages which have been partially studied transformationally, it turns out that all transformations are either unary, i.e., operate on one sentence to produce a sentence, or else binary, i.e., operate on two sentences to produce a single sentence. Every sentence of the language is thus decomposed by (the inverse of) a binary transformation into two sentences (with a binary operator on them), or else by (the inverse of) a unary transformation into one sentence with a unary operator on it. Each sentence of these decomposition products is in turn decomposable either into two sentences with a binary on them or one sentence with a unary on it. And so on till we reach the elementary sentences. Each sentence of the language can therefore be represented in a unique way (except, in some cases, for order) by a sequence of elementary sentences with unary and binary operators on them and on the operators. Thus we represent (avoiding compact symbolisms): *The book was taken by a man whom he and I know.*

1. *A man took the book.*
2. Passive (1)
3. *He knows a man.*
4. *I know a man.*
5. *and* (3, 4), with repeated corresponding words zeroed.
6. *wh* (2, 5) with respect to *man* (more exactly: with respect to the first N of 1 and the last N of 3, 4).

This representation supplies a normal form for every sentence, that is, a fixed form in which every sentence can be written; and I should explain that since transformational methods are entirely formal, i.e., based on the combinations of words as members of classes, and not based upon meanings, the obtaining of the normal form can be done mechanically. In fact, it can be done in principle by a computer, and in principle for all sentences of the language (although there are problems of idiomatic expressions and the like). The transformational rules decompose one form into another. The computer can be programmed to recognize forms as sequences of classes (possibly with the omission of something that is expected as part of the sentence), plus stated constants such as *is* and *-en* and *by* of the passive. On this basis the computer can rearrange the words of the sentence into the form of the decomposition-product sentence, and so down to the elementary sentences (which, because of an algebraic property which they have, I have called the kernel sentences) and the unary or binary operations on them.

The normal form opens the way to a large number of applications. For instance, suppose that in a scientific article every sentence in the article is decomposed in this way. (This is something which we have actually carried out.) Then we shall find that certain kinds of information (e.g., about the events that are being studied) are contained in the kernel sentences, certain kinds of information (e.g., about the analyses and activities of the scientists) are contained in a certain set of meaning-carrying unary operators on the kernel sentences, and certain kinds of information (e.g., quasi-logical relations) are contained in a different set of unary operators and in the binary operators.

Furthermore, if two words are informationally related to each other, that is, if the article speaks about some connection between their concepts, the two words are necessarily to be found inside of one kernel sentence (plus certain adjuncts) or in

two kernel sentences which are connected to each other by certain chains of binary operations. A statement of this kind cannot be made about unanalyzed sentences, for two words can occur in one sentence, even next to each other, without having any contentual relation; and two concepts may be related in a sentence without one of them being expressed by any word in the sentence (e.g., if the word has been dropped owing to a zeroing transformation).

I should say that, in addition to this, there is a mathematical character to this theory, because what we have here is a set of objects, sentences, and an operation on the set itself, an operation which decomposes or composes the objects in this set into other objects of this set, either one or else two at a time. What is of special interest here is that the objects on which this and other mathematically defined operations act are objects of the real world, the set of sentences of a language.

There is one fact which has to be stated, in addition to the existence of the transformational relation. This is that the transformations of a language form a tightly knit structure. It is not the case that each language simply has some arbitrary permutations and constant-addings which create a new form out of an old form, in such a way that the sentences of the old form are to be found also (as transforms) in the new form. Rather, it turns out that in each language there is a very small set of operations on sentences, which satisfy the definition of a transformation and which have an understandable, often informationally or grammatically functional, character. The hundred or so major transformations that a language has are each a particular application or succession of some of these elementary transformations.

I shall now give a sketch of the elementary sentence structures and transformations for English. The kernel structures are primarily (each with tense before the verb; I disregard here certain problems of *the,* etc.) :

N V	*A tree fell.*
N V N	*A rock struck a tree.*
N V P N	*A child relies on luck.*
N V N P N	*The man attributed the picture to Vasari.*
N be A	*The man is glad.*
N be P N	*The box is near a corner.*
N be $D_{loc.}$	*The man is here.*
N be N	*Man is a mammal.*

The elementary transformations are:

1. Adjuncts (in effect, modifiers) to the parts of a kernel sentence or to the whole sentence. E.g.:

to *N*:	*a, only,* etc.	(*a man, men only*)
to *A*:	*very, quite,* etc.	(*very large*)
to *V*:	*quite, just,* etc.	(*I just forgot*)
to *K* (kernel sentence):	*however,* etc.	

2. Sentence operators. These are special verbs (with their subjects or their objects), or adjectives or nouns, which operate on a sentence by making it their object or their subject. E.g.:

> *I know that he came.*
> *I wonder whether he came.*
> *I know of his coming.*
> *That he came surprised me.*
> *That he came is clear.*
> *That he came is a fact.*
> *His coming is clear.*

Each of these can operate on any sentence of the language, including their own resultants:

> *He suspects that I know that he came.*

There are restricted forms of these operators in which the subject of the operator and the subject of the included sentence are necessarily identical (the second usually zeroed):

> *He does (his) studying at night.*
> *He tried to come.*
> *He began to come.*

There are also several major groups of adverbs which appear as adjuncts of the verb, but also as operators on the sentence:

> *He speaks slowly.*
> *His speaking is slow.*

3. Connectives between two sentences:

coordinate (*and, or, but*) between any two sentences,

comparative (*than, as*) between two sentences the first of which contains a comparative marker (*more, less,* etc.);

subordinate (*because, while, after,* etc.) between any two sentences;

wh-words between two sentences which contain an identical noun; the common *N* is omitted from the second sentence, and the *wh* plus second sentence becomes an adjunct of the common *N* in the first:

> [*The book was poor*] *wh*- [*He read the book*]
> →*The book which he read was poor.*

4. Under specific conditions, zeroing of words which can be determined from other words (or from other occurrences of the same word) in the sentence: Compare *I want you to go* with *I want to go,* where obviously the subject of the second verb is zeroed when it is the same as the subject of the first.

5. There are certain analogic extensions of these elementary transformations, which produce forms like those produced above but on slightly different subclasses of words.

6. There are inverses of all these transformations, which take a sentence that looks like (but is not) the resultant of one of these transformations and creates a pseudo-original for it. This is a frequent event, and is similar to what has been called back-formation in linguistics. It is an unexpected result that many transformations, such as *The writing of letters was by him, His writing was of letters,* from *He wrote letters,* are simply inverses of the elementary transformations listed above. Others, such as *A letter is what he wrote, It is letters that he wrote, What did he write?* are successions of the above.

7. Finally there are certain permutations which create peculiar sentences in which the syntactic character of the parts are unchanged even though their position is changed. Thus in *This I like,* the word *this* is still the object of the verb even though it is in first position. So also for *this* in *This say the scientists.*

All the elementary transformations (and therefore all the complex transformations which are built out of them), aside from this last group, form sentences whose structure is similar to the kernel structures except for a limited number of additions and changes which are specified when we list the transformations. Each transformation acts on particular structures (of the kernel, and resultants of particular transformations) and produces out of them a particular structure. Thus one transformation can act on the resultant of another, if the earlier one has produced a structure which matches the operand of the later one as to constants, arrangement of word-classes, and the subclasses involved. The complicated transformations, such as the question, are simply successive applications of these elementary transformations, matching resultant and operand. It is the detail of matching subclasses, of limitations as to which words take which affixes, and the like, that makes language so complicated. The essential structure is simply that which has been sketched here.