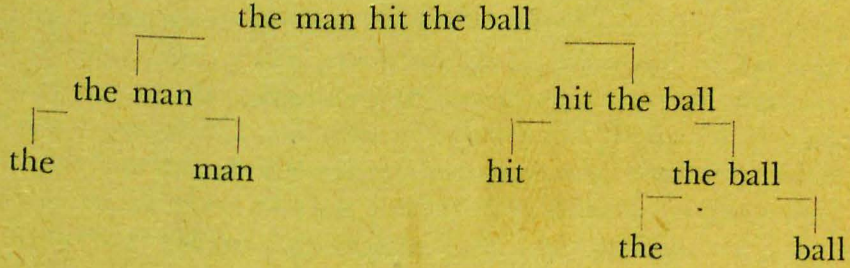
# SYNTACTICAL CATEGORIES[1]

## I. Immediate Constituent Analysis

Modern structural linguists are used to describe the syntactic structure of *(well-formed)* sentences and *phrases*[2] by parsing them "into two or more contiguous constituents, either of which is already a final constituent or else is itself parsible into two or more immediate constituents, etc.»[3] Consider for instance the sentence

(1) the man hit the ball

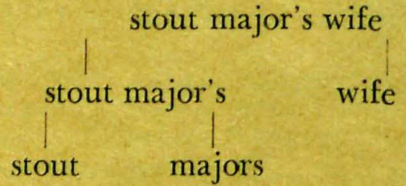
which is parsible in the following way :[4]



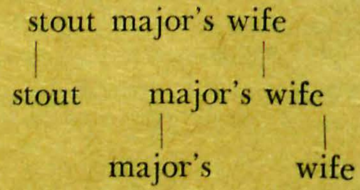

As shown in this diagram, sentence (1) is divided into two IMMEDIATE CONSTITUENTS — ICs for short — *'the man'* and *'hit the ball'*. In their turn, the first is divided into two ICs *'the'* and *'man'*, while the second is divided also into two ICs *'hit'* and *'the ball'*. Finally *'the ball'* is divided into two ICs *'the'* and *'ball'*. Now the ICs of the ICs of a given phrase are called its CONSTITUENTS OF THE SECOND ORDER, the ICs of the ICs of the ICs of the phrase are called its CONSTITUENTS OF THE THIRD ORDER, etc.[5] Thus (1)'s constituents of the second order are *'the'* (first occurence), *'man'*, *'hit'*, *'the ball'*; while the same sentence's constituents of the third order are *'the'* (second occurence) and *'ball'*. (1) has no constituents of an order higher than the third, so that (1) is itself called a phrase OF THE THIRD ORDER.[6] The ICs of a given phrase, together with its constituents of the second, third, ... order constitute the phrase's CONSTITUENTS[7]. Constituents which cannot be parsed into two or more ICs are called ULTIMATE CONSTITUENTS[8] Hence, the ultimate constituents of a phrase consist in the sundry words by means of which it is built up. A one-word phrase has only one ultimate constituent — itself.

The parsing of a given phrase into its various constituents, up to the ultimate ones, is called the IMMEDIATE CONSTITUENT ANALYSIS of that phrase. Such an analysis is not necessarily unique. For example the phrase *'stout major's wife'* (in such a context as *'he was dancing with the stout major's wife')* can be parsed in two different ways :[9]

(i)




(ii)




Thus *'stout major's'* is a constituent in case (i) but not in case (ii), while *'major's wife'* is a constituent in case (ii) but not in case (i). Such an *ambiguity* (resulting from the *possibility of parsing a phrase in two or more different ways)* is called CONSTRUCTIONAL HOMONYMITY.[10]

Constructional homonymity is clearly a kind of *syntactical ambiguity*. Non the less, it can reflect a genuine *semantical* ambiguity, either. Thus *'stout major's wife'* refers in case (i) to the wife of a stout major, and in case (ii), to the stout wife of a major. There are, however, also cases of constructional homonymity which are not reflecting any semantical ambiguity at all. E.g., sentence (1), besides its customary parsing mentioned above, can be parsed also in the following way :

```
                the man hit the ball
               |           |       |
         the man          hit     the ball
        |       |                |        |
       the     man              the      ball
```

We see that *'hit the ball'*, although being an IC of (1) with respect to the first parsing, is, with respect to the latter, no constituent of (1) at all. But these two parsings, although radically different from the syntactic point of view, are not bound to any difference in reference, or else in meaning.[11]

*'Stout major's wife'* is an instance of a phrase, each (continuous) part of which is a constituent. But such phrases are exceptional. In general, a phrase has parts which are not its constituents. E.g., *'hit the'* is not a constituent of *'the man hit the ball'*. Consequently, we see that immediate constituent analysis is not an automatic task. It will also not help to construe the constituents of a given phrase as those parts whose meanings constitute the meaning of the whole phrase. Indeed, the best way for determining whether the meaning of a given phrase is composed of the meanings of certain of its parts, consists in inquiring as to whether these parts are — or are not — constituents of the phrase in question. Hence immediate constituent analysis has to be explained from a purely *syntactic* point of view (rather than from a *semantic* one).

In Section III we shall expose a syntactic method yielding an effective (operational) decision procedure for determining *(a)* whether a given expression is *well-formed;* and *(b)* whether a part of a phrase is a *constituent,* or else an *IC,* of this phrase. The same method yields also a systematic and exhaustive *classification of all the well-formed expressions* (phrases) of any language to which the method can be applied.

## *II. Operators and Operands*

Let us consider a simple noun-verb sentence such as

(2) John runs.

From the point of view of Aristotelian logic (2) is considered to be short for *'John is running'*, and it is construed to mean that the predicate *'running'* is predicated of the subject *'John'*. In modern nominalistic terms, we interpret (2) to mean rather that the predicate (i.e. the verbal phrase) *'runs'* applies to John. In both interpretations the verbal phrase is not placed on the same footing as the nominal phrase, i.e. the subject. The predicate is rather considered as *determining, governing* or *operating on* the subject. So the subject is regarded as a *determined, governed, dependent* phrase. We shall say that the predicate is an OPERATOR operating on the subject which shall therefore be called an OPERAND[12]. Thus *'runs'* is the operator of sentence (2), while *'John'* is its operand.

Consider now the sentence

(3) John loves Mary.

Here the verbal phrase *'loves'* consists in a transitive verb and applies consequently to two entities, viz. John and Mary. So *'loves'* operates on two nominal phrases *'John'* and *'Mary'*. Hence (3) has one operator and two operands.

We shall now make the following *basic assumption :*

> *Every phrase of more than one word is divisible (exhaustively) into two or more (continuous and non-overlapping) ICs, such that just one of these ICs is an operator determining each one of the remaining ICs — the operands.*[13]

If A is the operator of a phrase B, B is called a CLOSURE of the operator A. Hence, a closure of a given operator is composed of the operator itself and of the operands that it determines. E.g., *'John runs'* is a closure of the operator *'runs'*; *'he runs'* is another closure of the same operator. *'John loves Mary'* is one closure of the operator *'loves'*, while *'I love you'* is another one.

The operands can stand to the left or to the right, or else, on both sides of the operator. We shall call LEFT OPERAND an operand which stands on the left-hand side of its operator, and RIGHT OPERAND an operand which stands on the right-hand side of its operator. For example, *'John'* is a left operand in *'John runs'*, but a right operand in *'little John'*. Indeed *'little John'* has for operator *'little'* and for operand *'John'*.

An operator having only right operands is called a PREFIXED OPERATOR, or PREFIXE, for short; while an operator having only left operators is called a SUFFIXED OPERATOR, or SUFFIXE, for short. On the other hand, we can classify operators according to the number of their operands. An operator having n operands is called an *n-place* (or *n-ary)* operator. 1 - place operators are called also UNARY and 2-place operators BINARY. A binary operator written between its two operands (i.e. an operator having just one left operand and one right operand) is called an INFIXED OPERATOR, or INFIXE for short.[14]

For instance, in *'John runs' 'runs'* is a suffixed unary operator, and *'little'* is in *'little John'* a prefixed unary operator. On the other hand, in *'John loves Mary' 'loves'* is a (binary) infixed operator.

Now every (English) phrase is — in some context — an operator or an operand. Indeed, any *well-formed* expression is (by definition) either a *sentence,* or else a *contituent* of some sentence. A constituent is necessarily (as a result of our basic assumption) either an operator or an operand, while any sentence is (in some context) an operand. Indeed, let S be an arbitrary (English) sentence. Then we can form the denial of S, viz. the expression "it is not the case that S". The latter has for operator *'it is not the case that'* and for operand the sentence S itself. Hence we see that *every phrase is either an operator or an operand.* On the other hand, it is obvious that only a well-formed expression can be a sentence or a constituent of a sentence. We arrive thus to the following result :

> *An expression (i.e. a sequence of words) is well-formed if and only if it is an operator or an operand.*

We shall attempt to divide all phrases (of any given language) into different *categories,* by way of taking into account merely the "operator-or-operand status" of each phrase in its various contexts. Such categories, being based exclusively on *syntactic* features, are called SYNTACTICAL CATEGORIES.[15] We shall use here for *'syntactical categories'* the abbreviation 'SCs'. Now although it is possible to divide all phrases of certain formalized languages into mutually exclusive SCs, that is not possible for the case of natural languages, because of the *syntactical ambiguity* of their phrases. E.g., certain English words such as *'show'*, *'work'*, *'thought'* are nouns in some contexts and verbs in some other contexts. But, as we shall see below, nouns and verbs constitute basically different SCs. The concept of a syntactical category would loose its sense if one unites nouns and verbs into a single category. Consequently, we must either renounce to apply the concept of syntactical categories to natural languages altogether, or else we must allow the assignement of *more than one* SC to certain phrases. We choose, following Y. Bar-Hillel,[16] the second alternative; so that instead of «dividing» all phrases of a given natural language — say of English — into different SCs, we shall rather «ascribe» each phrase to one or more pre-established SCs.

We can divide all expressions (whether well-formed or not) into the following, mutually exclusive classes :

(a) Expressions which are both operands in some context and operators in some other context.

(b) Expressions which are operands in some context but not operators in any context.

(c) Expressions which are operators in some context but not operands in any context.

(d) Expressions which are neither operands nor operators in any context.

As shown above, all and only those expressions which are well-formed are operators or operands in some context. Hence the class of well-formed expressions, i.e. of phrases, consists in the union of the classes (a), (b) and (c). No expression belonging to (d) is a phrase.

In case of a natural language such as English, the class (b) is not empty; it includes at least the *sentences* of the language. Indeed sentences are *operands* in certain contexts, but they are obviously not operators in any context. Most of *nouns* and, in general, of *nominals* (i.e., name-like phrases) belong also to class (b). Syntactically ambiguous nouns such as the above-mentioned words *'show'*, *'work'*, *'thought'* etc. do not belong to (b), since they are verbs and thus operators in certain contexts.

Let us call *fundamental phrases* the elements of class (b). So sentences and (most of) nominals are fundamental phrases. It is possible to distinguish various kinds of *sentences* such as *declarative, interrogative, imperative, optative, exclamatory* ones; as well as various kinds of *nominals* such as *singular* and *plural, concrete* and *abstract, first person, second person* and *third person, proper names, common nouns,* etc. However, for the sake of simplicity, we shall assume only two categories of fundamental phrases, viz. the SC of (declarative) SENTENCES (*s* for short) and the SC of NOMINALS (*n* for short), disregarding non-declarative sentences and gathering all kinds of nominals together. The SCs assigned to fundamental phrases *qua* fundamental ones are called FUNDAMENTAL SCs. Our assumption can then be formulated as follows : There are just *two* different fundamental SCs, the SC of *sentences,* i.e. *s,* and the SC of *nominals,* i.e. *n.*

We shall call *operator phrases* all phrases belonging to one of the classes (a) and (c). Hence, while a "fundamental phrase" is one which is in some context an operand, but not an operator in any context; an "operator phrase" is one which is in some context an operator. Consequently, words like *'show', 'work', 'thought'* (being operators in certain contexts) are operator phrases and not fundamental phrases, although, in their quality of nouns, they are ascribed to the fundamental SC *n.*

All operator phrases are ascribed — in their quality of operators — to non-fundamental SCs. The latter are called OPERATOR SCs. Every SC is either a *fundamental SC* (i.e., is identical with *s* or with *n*) or else is an *operator SC.* No fundamental SC is an operator SC and no operator SC is a fundamental SC. However the same phrase can be ascribed both to a fundamental SC and to an operator SC. That is the case, in particular, for nouns (like *'show', 'work', 'thought',..*) which are not fundamental phrases.

As we have seen above, the reason for admitting just two distinct fundamental SCs is rather conventional. We could as well admit any other number (though *two* seems to be the minimum). On the other hand, once certain fundamental SCs. are admitted, the operator SCs are strictly determined. Such a determination is even independent of the fact of there being any phrases at all to which these SCs built up *in abstracto* could be assigned.

The whole hierarchy of *operator SCs* can be constructed by means of the following two basic rules :

(I) i. *An operator and any of its operands belong to different SCs.*[17]

ii. *An operator and any of its closures belong to different SCs.*

For example, *'runs'* being an operator determining *'John'*, *'runs'* and *'John'* belong to different SCs. On the other hand, *'John runs'* being a closure of *'runs'*, *'runs'* and *'John runs'* belong to different SCs.

(II) *Any two operator phrases A and B belong to the same SC if, and only if, there is a closure A' of A and a closure B' of B such that :*

(a) *A' and B' contain the same number of left and right operands;*

(b) *the corresponding operands* — i.e. the *i*-th left operands or the *i-th* right operands ($i = 1, 2, 3,\ldots$) in A' and B' respectively — *belong to the same SC;*

(c) *the closures A' and B' belong to the same SC.*

For instance, we can show by means of (II) that *'loves'* and *'hit'* belong to the same SC. Indeed there is a closure of *'loves'* say *'John loves Mary'* and a closure of *'hit'* say *'the man hit the ball'* such that :

*(a)* Each closure has just one left and one right operand.

*(b)* The left operands (viz. *'John'* and *'the man'* respectively) as well as the right operands (viz. *'Mary'* and *'the ball'* respectively) belong to the same SC, namely to the category *n*.

*(c)* The two closures belong to the same SC, viz. the category *s*.

## III. *Quasi-Arithmetical Notation*

Let us use Latin capitals as variables ranging over the expressions of the object-language, and lower case Greek characters as variables ranging over the SCs. We shall say then that a sequence $\alpha_1,\ldots,\alpha_n$ of SCs is a SC-SEQUENCE *of* an expression Y whenever there are expressions $X_1,\ldots,X_n$ such that $Y = X_1\ldots X_n$ (i.e. Y is the concatenation of $X_1,\ldots X_n$) and $X_1,\ldots,X_n$ belong respectively to $\alpha_1,\ldots,\alpha_n$. (Every expression, being a concatenation of words, has at least one SC-sequence.)

Consider a phrase C such that $C = B_m \ldots B_1 D A_1 \ldots A_n$ and $A_1,\ldots,A_n$, $B_1,\ldots,B_m$,C belong respectively to the SCs $\alpha_1,\ldots\alpha_n$ $\beta_1,\ldots,\beta_m,\gamma$. Let D be the operator of C. Then we shall designate any SC assigned to the operator phrase *qua* operator in C by means of symbolic expressions of the form '$[\beta_m,\ldots,\beta_1\backslash\gamma/\alpha_1,\ldots,\alpha_n]_k$'. The subscript '*k*' is used here to distinguish (if any) the different SCs which can be assigned to D.

Now let C' be any phrase (possibly identical with C) such that D' is the operator of C' and the corresponding operands of C and C' belong to the same SCs (or are even identical). Then, in virtue of rule (II) (of section II), any one of the SCs assigned to D' (on the basis of its being the operator of C') is identical with some SC assigned to D (on the basis of the latter's being the operator of C). I.e., for all k and h

$$[\beta_m,\ldots,\beta_1\backslash\gamma/\alpha_1,\ldots,\alpha_n]_k = [\beta_m,\ldots,\beta_1\backslash\gamma/\alpha_1,\ldots,\alpha_n]_h$$

so that the SC assigned to any operator phrase D, in virtue of its being the operator of a phrase C, is uniquely determined by the SCs of the operands of C and by the SC of C itself. We can consequently drop the subscript 'k' and designate the SC assigned to D (*qua* operator of C) by[18]

(4) $$[\beta_m,\ldots,\beta_1\backslash\gamma/\alpha_1,\ldots,\alpha_n]$$

We can say then that

(5) $$\beta_m,\ldots,\beta_1,[\beta_m,\ldots,\beta_1\backslash\gamma/\alpha_1,\ldots,\alpha_n],\ \alpha_1,\ldots,\alpha_n$$

is a SC-sequence of the phrase

$B_m \ldots B_1 D A_1 \ldots A_n$

which belongs to the SC $\gamma$. We shall show, furthermore, that *every* expression which has (5) as a SC-sequence belongs to the SC $\gamma$. Indeed, let U be any such expression. Then U will be a concatenation of the form

$Y_m \ldots Y_1 Z X_1 \ldots X_n$

such that $X_1\ldots,X_n$, $Y_1,\ldots,Y_n$,Z belong respectively to the SCs $\alpha_1,\ldots,\alpha_n,\beta_1,\ldots,\beta_m$ and $[\beta_m,\ldots,\beta_1\backslash\gamma/\alpha_1,\ldots,\alpha_n]$. It follows that Z is the operator of U and U belongs to the SC $\gamma$. We express this property by the CANCELLATION RULE.[19] *(5)* $\rightarrow \gamma$, i.e.

$$\beta_m,\ldots,\beta_1,\ [\beta_m,\ldots,\beta_1\backslash\gamma/\alpha_1,\ldots,\alpha_n],\ \alpha_1,\ldots,\alpha_n \rightarrow \gamma$$

which is to be read *"the SC-sequence (5) DIRECTLY CANCELS to $\gamma$".*

We shall say, furthermore, that a SC-sequence *a* CANCELS to a SC-sequence (or, as a particular case, to a SC) *b*, in case *b* results from *a* by finitely many applications of the cancellation rule (more exactly, if there exist SC-sequences $c_1,\ldots,c_n$ such that $c_1 = a$, $c_n = b$ and $c_i$ directly cancels to $c_{i+1}$ for $i = 1, 2, \ldots, n-1$).

The import of the above mentioned rule (II) can be formulated in the following way :

$[\beta_m,\ldots,\beta_1\backslash\gamma/\alpha_1,\ldots,\alpha_n] = [\beta_h',\ldots,\beta_1'\backslash\gamma'/\alpha_1',\ldots,\alpha_k']$ with $m + n > 0$ and $k + h > 0$, if and only if $m = n$, $h = k$ and $\alpha_1 = \alpha_1',\ldots,\alpha_h = \alpha_h'$, $\beta_1 = \beta_1',\ldots,\beta_m = \beta_m'$, $\gamma = \gamma'$.

On the other hand, we can formulate the import of rule (I) (of section II) as follows :

(i) $[\beta_m,\ldots,\beta_1\backslash\gamma/\alpha_1,\ldots,\alpha_n] \pm \alpha_i \quad (i = 1,\ldots,n)$

$[\beta_m,\ldots,\beta_1\backslash\gamma/\alpha_1,\ldots,\alpha_n] \pm \beta_j \quad (j = 1,\ldots,m)$

(ii) $[\beta_m,\ldots,\beta_1\backslash\gamma/\alpha_1,\ldots,\alpha_n] \pm \gamma$

We can now define inductively the whole set of SCs in the following way :

(i) *s* and *n* are SCs.

(ii) If $\alpha_1,\ldots,\alpha_n,\beta_1,\ldots,\beta_m,\gamma$ are SCs, then $[\beta_m,\ldots,\beta_1\backslash\gamma/\alpha_1,\ldots,\alpha_n]$ is a SC.

(iii) There are no SCs other than those determined by (i) and (ii).

We shall call the symbolic expressions for SCs of the form '$[\beta_m,\ldots,\beta_1\backslash\gamma/\alpha_1\ldots,\alpha_n]$' (with $m + n > 0$) *QUASI-FRACTIONS,* '$\gamma$' being the NUMERATOR; '$\beta_j$' ($i = 1,\ldots,m$) the *i-th LEFT DENOMINATOR* and '$\alpha_i$' ($i = 1,\ldots,h$) the *i-th RIGHT DENOMINATOR*. There is a *one-one correspondence* between the set of operator SCs and the set of quasi-fractions. Indeed every operator SC is (in virtue of the above mentioned inductive definition) designated by a quasi-fraction, while two distinct quasi-fractions designate always different operator SCs. On the other hand, no quasi-fraction designates a fundamental SC. Calling *category symbol* any quasi-fraction or any one of the letters *'s'* and *'n'*, we see that there is a one-one correspondence between the set of SCs and the set of category symbols.

Having thus constructed in *abstracto* the whole hierarchy of operator SCs, we shall now assume the task mentioned in the last paragraphe of section I.

As Bar-Hillel put it, the theory of SCs is "not... a method which a linguist might use to *arrive* at an analysis of a linguistic corpus, but only... a new way in which he could *present* the results of his investigations.[20] We assume that these results are embodied in the establishment of a "SC-DICTIONARY",[21] i.e. a list ascribing each word (of the vocabulary of the language under investigation — English in our case) to a finite number of SCs.

The preparation of such a SC-dictionary proceeds in a step-by-step way :

(i) We know first (say on the basis of our grammatical intuition) certain words to be nominals in the context of certain expressions which we know to be sentences. We can thus assign to these words and expressions, respectively the SCs $n$ and $s$.

(ii) Every sentence has just one operator (among its ICs). Hence if all words but one of a given sentence are nominals (in that context), we know that the remaining word is the operator. We assign to this word a SC of form $[n\ldots n\backslash s/n\ldots n]$.

Suppose e.g. that we know *'John runs'* to be a sentence with *'John'* as a nominal. Then (since a nominal cannot, *qua* nominal, be an operator) we can infer that *'runs'* is the operator, ascribing it, consequently, to $[n\backslash s]$. Similarly, knowing that *'John loves Mary'* is a sentence, while *'John'* and *'Mary'* are nominals in this context, we conclude that *'loves'* is an operator belonging to $[n\backslash s/n]$.

Operators belonging to SCs of the form $[n\ldots n\backslash s/n\ldots n]$ are *predicates.* They are called also *predicators* or VERBALS. Not all verbals are single-word expressions. E.g., in *'John is running'* the verbal consists in the phrase *'is running'* (and not in the single word *'running'* alone). In order to ascertain the operator character of *'is running'*, it is clearly not sufficient to know merely that *'John'* is a nominal and *'John is running'* a sentence; one must also know that *'John'* is the *sole* nominal and, what is more, the sole *operand* occuring in the sentence in question. This we assume to know by means of our grammatical intuition. In the same way, knowing that *'John'* and *'Mary'* are nominals constituting the sole operands of the sentence *'John has been loving Mary'*, we can reach the conclusion that *'has been loving'* is a verbal (belonging to $[n \backslash s / n]$ ).

(iii) By means of verbals, we can then ascertain complex nominals as the many-word operands of these verbals. For instance, once we recognize *'are running'* as a verbal (say as a result of its being the operator of *'boys are running'*), we can infer that *'small boys'* is a nominal in the context of the sentence *'small boys are running'*. Similarly, we can show that *'the boys'*, *'some boys'*, *'all boys'*, *'very small boys'*, *'the very small boys'*, etc. are all nominals. We can also show (say with the help of the verbal *'is running'*) that *'the boy'*, *'a boy'*, *'some boy'*, *'any boy'*, *'every boy'*, *'the small boy'*, *'the very small boy'*, *'the very small and thin boy'*, etc. are nominals too.

We can show also, in the same way, that *pronouns* are to be ascribed to the SC of nominals. E.g., *'I'* is operand in *'I did'*, *'he'* in *'he runs'*, *'her'* in *'he loves her'*, etc.

(iv) The complex nominals thus obtained are usually composed of a unary operator with a nominal for operand: *'small boys'* is constituted by the operator *'small'* and the operand *'boys'*, *'very small boy'* of the operator *'very small'* and the operand *'boys'*. Consider e.g., *'small boys'*. We know that this expression is a nominal, and we know that *'boys'* is a nominal. Hence the remaining word *'small'* must be the operator of *'small boys'*. Hence, *'small'* can be ascribed to $[n/n]$. Such operators are called ADJECTIVALS[22] (or more precisely, *"attributive adjectivals"*). English adjectivals are usually prefixed operators. An exception is constituted by the adjectival *'general'* which (although being in most contexts also a prefixed operator) is in certain cases a suffixed operator, e.g. in *'inspector general'*, *'attorney general'*, etc. In these contexts *'general'* belongs to $[n\backslash n]$.

While *'small'* is a single-word adjectival, *'very small'* is a complex one. *'Very small boys'* being a nominal and *'boys'* a nominal which is the sole operand in this context, we infer indeed that *'very small'* is an operator belonging to $[n/n]$.

(v) Consider a complex adjectival like *'very small'*. *'Small'* being also an adjectival, viz. one which is the (sole) operand in *'very small'*, we conclude that *'very'* is an operator belonging to the SC $[[n/n]/[n/n]]$. We call such operators (which are *modifiers* of adjectivals) AD-ADJECTIVALS. We can similarly obtain a great variety of ad-adjectivals by means of complex adjectivals.

(vi) Consider the sentence *'John runs quickly'*. *'John'* being a nominal and, furthermore, the sole operand in this sentence, we see that *'runs quickly'* is a complex verbal (belonging to $[n\backslash s]$). Suppose we know that the verbal *'runs'* is the operand. Then we infer that *'quickly'* is an operator belonging to $[[n\backslash s] \backslash [n\backslash s]]$. Such operators are called ADVERBALS.

Continuing in this way, we can — in principle — assign to each word one or more SCs, establishing thus a fictitious *SC-dictionary* of English.

Once that such a "distionary" has been established, it becomes possible to determine in a mechanical, "quasi-arithmetical" way whether any given expression (i.e. word sequence) is a *sentence*, or in general a phrase and, if a phrase, what its constituents are.

We have now reached a level of linguistic analysis in which we can abstract from all our grammatical intuition, assuming no other knowledge than that of the SC-dictionary, together with the rules for dealing with SCs. At this level, a *well-formed expression* (or *phrase*) is defined to be any word sequence having at least one SC-sequence which cancels to some SC. In particular, a *sentence* is any word sequence having at least one SC-sequence which cancels to *s*. We assign to each well-formed expression all and only those SCs to which its SC-sequences cancel.

SC-sequences which cancel to some SC are called CONNEX SC-SEQUENCES. Calling CONNEX EXPRESSION any expression having at least one connex SC-sequence, we can then define a *well-formed expression* simply as being a connex expression. SC-sequences and expressions which are not connex are called DISCONNEX. The cancellation rule for quasi-fractions yields clearly a procedure for testing the CONNEXITY of any given SC-sequence and, consequently, of any expression (i.e. word sequence).

Consider e.g., the word sequence *'John runs'*. We suppose to have found out in the SC-dictionary that *'John'* is ascribed to $n$ and *'runs'* to $[n\backslash s]$. We infer thus that *'John runs'* can be correlated with the SC-sequence

$$n,\ [n\backslash s].$$

This SC-sequence cancels to $s$, for according to the cancellation rule,

$$n,\ [n\backslash s] \mapsto s.$$

We can conclude then that *'John runs'* is a *sentence.*

Let us now examine, as a more complex example, the expression (1) mentioned in section (I) (i.e., *'the man hit the ball')*. Suppose that the SC-dictionary ascribes to the words composing this expression the following SCs (listed respectively below each word) :[23]

| the | man | hit | the | ball |
|---|---|---|---|---|
| $[n/n]$ | $n$ | $[n\backslash s/n]$ | $[n/n]$ | $n$ |
| | $[n\backslash s/n]$ | $[[n\backslash s]/n]$ | | $[n\backslash s]$ |
| | | | | $[n\backslash s/n]$ |

When *'hit'* is ascribed to $[n\backslash s/n]$, any SC-sequence of (1) will cancel to a SC (viz. to $s$) if and only if the SC-sequences of *'the man'* and *'the ball'* will both cancel to $n$. Among the SC-sequences of *'the man'* "$[n/n]$, $[n\backslash s/n]$" does not cancel to $n$ (nor to any other SC), while among the SC-sequences of *'the ball'* neither "$[n/n]$, $[n\backslash s]$" nor "$[n/n]$, $[n\backslash s/n]$" are cancelling to $n$. Exactly the same conditions are required for the case *'hit'* is ascribed to $[[n\backslash s]/n]$. Consequently, the *twelve* possible SC-sequences which can be correlated with (1) reduce to *two,* viz. :

(i) $[n/n],\ n,\ [n\backslash s/n],\ [n/n],\ n$

(ii) $[n/n],\ n,\ [[n\backslash s]/n],\ [n/n],\ n$

It follows thus, that in the context of *'the man hit the ball'*, the SC $[n\backslash s/n]$ cannot be assigned to *'man'*, while the SCs $[n\backslash s]$ and $[n\backslash s/n]$ cannot be assigned to *'ball'*. In this way, context reduces the *syntactical ambiguity* resulting from the assignment of more than one SC to words.

Now both SC-sequences cancel to $s$, respectively by way of the following DERIVATIONS :

(i) $[n/n],\quad n,\quad [n\backslash s/n],\quad [n/n],\quad n$

$n,\quad [n\backslash s/n],\quad n$

$s$

(ii) $[n/n],\quad n,\quad [\,[n\backslash s]/n],\quad [n/n],\quad n$

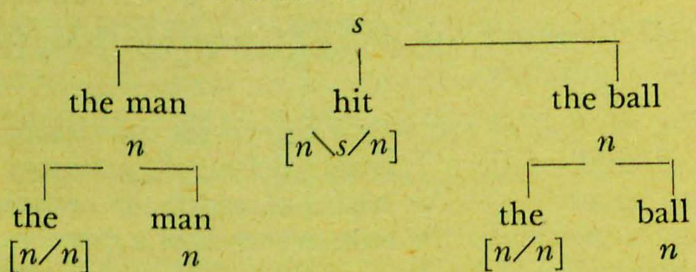
$n,\quad [\,[n\backslash s]/n],\quad n$

$n\quad [n\backslash s]$

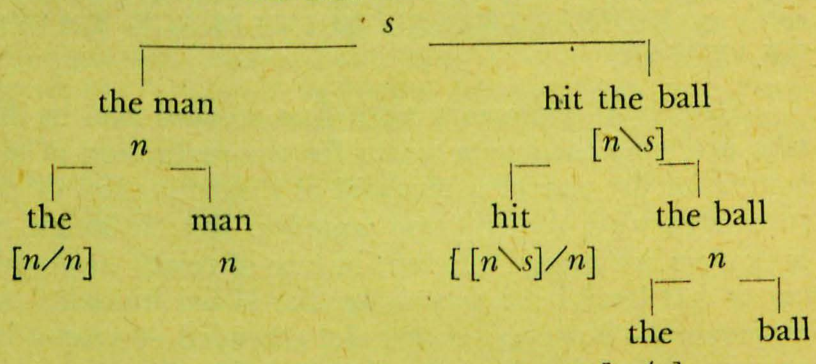
$s$

It follows, that (1) is in both cases a sentence. The constituents of this sentence are given by the following two diagrams called the TREE EXPANSIONS of the sentence :

(i) the man hit the ball




(ii) the man hit the ball




The frame of SCs underlying a tree expansion of a given phrase describes the CONSTITUENT STRUCTURE of that phrase. A phrase which has two or more different constituent structures is called CONSTRUCTIONALLY HOMONYMOUS. The sentence (1) is thus contructionally homonymous since it has two different constituent structures, viz. those determined by the tree expansions (i) and (ii) respectively.

The constructional homonymity of *'the man hit the ball'* is the result of its being correlated with *two* different SC-sequences. Constructional homonymity can, however, result also from a *single* SC-sequence. That would be the case for a SC-sequence yielding two or more different constituent structures. Consider e.g., the expression *'Paul thought that John slept soundly'*,[24] assuming that it can be "categorized" in the following way (by means of the "SC-dictionary") :

| Paul | thought | that | John | slept | soundly |
|---|---|---|---|---|---|
| $n$ | $[[n\backslash s]/n]$ | $[n/s]$ | $n$ | $[n\backslash s]$ | $[[n\backslash s]\backslash[n\backslash s]]$ |

It can easily be shown that such a SC-sequence yields two different constituent structures.

The method of syntactical categories constitutes a GRAMMAR. Indeed, modern linguists define a *"grammar"* of a language to be any device by means of which the constituent structure, and in particular the sentencehood, of a given expression of that language could be determined.[25] The grammars constituted by the method of syntactical categories are called CATEGORIAL GRAMMARS. The particular method we have used here constitutes a BIDIRECTIONAL MANY-PLACE CATEGORIAL GRAMMAR.

Categorial grammars were first worked out by Bar-Hillel, who proposed them as a means for the realization of *mechanical translation* by way of a mechanical determination of the constituent structure of any given sentence. But although categorial grammars are indeed efficient in case of relatively simple sentences, it has later been shown, by Bar-Hillel himself,[26] that these grammars — however refined and amended — cannot be used as a practical tool for the determination of the constituent structure of *all* sentences (of a natural language). Now, a *categorial grammar* constitutes a particular formalization of the customary *"immediate constituent model"* used by modern structural linguists, so that the inadequacy of the former reflects the inadequacy of the latter. This inadequacy has indeed been recognized since the advent of the *"transformational models"*. But the transformational models are not suppressing but merely supplementing the immediate constituent model. The latter "remains intact for a certain kind of simple sentences, the so-called *kernel sentences* (or rather for their underlying *terminal strings*" — and Bar-Hillel's "method of mechanical structure determination remains therefore valid for these sentences — but has to be supplemented by additional procedure, the so-called *transformations,* in order to account for the synthesis of *all* sentences".[27]

*Teo Grünberg*

---

[1] This paper is based mainly on Yehoshua Bar-Hillel's recently published book *"Language and Information"*. (A substantial review — in Turkish — of this book by H. Batuhan has been published in *"Felsefe Arkivi"*, 15.) Prof. Bar-Hillel is nowadays the greatest authority on the subject of *syntactical categories*, especially with respect to their application to the syntax of natural languages. Bar-Hillel was fascinated by the topic of syntactical categories since the beginning of his career. Indeed his doctoral thesis has the title of *"Theory of Syntactical Categories"* (1947). Since then, he published a number of important articles on this subject: *"On Syntactical Categories"* (1950), *"A Quasi-Arithmetical Notation for Syntactic Description"* (1953), *"Some Linguistic Obstacles to Machine Translation"* (1960), *"On Categorial and Phrase Structure Grammars"* (1960, in collaboration with C. Gaifman and E. Shamir), *"The Role of Grammatical Models in Machine Translation"* (1962). All these papers (together with many others on various topics) are reprinted in *"Language and Information"* (Addison-Wesley Publishing Co., 1964).

[2] I use the term *'well-formed'* for denoting any grammatically correct expression. Instead of *'well-formed expression'* I say more often *'phrase'*. In particular, *words* as well as *sentences* are phrases.

[3] *See* Bar-Hillel, *"Language and Information"*, p. 188.

[4] *Cf.* N. Chomsky, *"Syntactic Structures"*, (Mouton and Co., The Hague, 1957) pp. 26, 27 ff.

[5] *Cf.* Bar-Hillel, *op. cit.*, p. 70.

[6] *Ibid.*, p. 70.

[7] *Cf.* Chomsky, *op. cit.*, p. 28.

[8] *See* C. F. Hockett, *"A Course in Modern Linguistics"*, (Macmillan, New York, 1958) p. 152. (We consider *words* as being the ultimate contituents, but we could also, with equal right, choose *morphemes* for the same purpose.)

[9] *See* Hockett, *op. cit.*, *pp.* 152-153.

[10] *Cf.* Bar-Hillel, op. cit., p. 81; Chomsky, *op. cit.*, p. 86.

[11] *Cf.* Bar-Hillel, *op. cit.*, pp. 70-71.

[12] *Cf.* Bar-Hillel, *op. cit.*, p. 193. Customarily the term *'argument'* is used instead of *'operand'*. Cf. e.g., I. M. Bochenski, *"On the Syntactical Categories"*, (in A. Menne (ed.), "Logico-Philosophical Studies", D. Reidel, Dordrecht-Holland) pp. 72 ff; Curry, *Foundations of Mathematical Logic*, (Mc Graw-Hill, 1963) p. 32.

[13] *Cf.* Bochenski, *op. cit.*, pp. 71-73; Bar-Hillel, *op. cit.*, pp. 65, 76, 188. The basic idea is due to K. Ajdukiewciz as exposed in his paper *«Die syntaktische Konnexitaet"* (1935). (The English translation of this article of capital importance will be published in the forthcoming volume of Mc. Call (ed.) "Polish Logic".)

[14] *Cf.* Curry, *op. cit.*, pp. 34-37. Instead of 'operator' Curry uses the term *'functor'*.

[15] The notion of *"syntactical categories"* originates with Husserl's "meaning categories" (Bedeutungskategorien). *Cf.* E. Husserl, *"Logische Untersuchungen"* (Max Niemeyer, Halle a.d.S., 1913) vol. II, part I, pp. 294-342; Bar-Hillel, *"Husserl's Conception of a Purely Logical Grammar"*, (Phil. and Phen. Res., 1957). The first precise theory of syntactical categories was formulated by S. Lesniewski (who called them "semantical categories"). *Cf. E. C.* Luschei, *"The Logical Systems of Lesniewski"; (North-Holland, 1962); M. Machover, *Contextual Determinacy in Lesniewski's Grammar"*, (Jerusalem, 1964). Lesniewski's theory was developed by Ajdukiewicz, and the latter's ideas matured with Bar-Hillel who applied them to the syntactic description of natural languages. *Cf.* also Bochenski, *op. cit.;* Bochenksi - Menne, *"Grundriss der Logistik"* (pp. 16, 17-18, 115-116); Curry and Feys, *"Combinatory Logic"*, (North-Holland, 1959) pp. 274-275; Fraenkel and Bar-Hillel, (North-Holland, 1958) *"Foundations of Set Theory"*, pp. 168-171; Curry, *op. cit.*, pp. 32-33.

Grammarians and linguists have also studied the subject : Jespersen used the very term of 'syntactical categories' in his *"Philosophy of Grammar"* (pp. 52-53); Harris studied the syntactical categories under the name of 'morpheme-sequence classes' *(vid.* Z. S. Harris, *"Methods in Structural Linguistics"*, esp. pp. 273ff) and Hockett under the name of 'form classes' (*vid.* Hockett, *op. cit.*, pp. 157 ff).

I myself have made much use of the concept of syntactical categories in my doctoral thesis (in Turkish) *"Anlam Kavramı üzerine bir deneme"* (An Essay concerning the Concept of Meaning) where I tried to apply them especially to semantical and ontological analysis. (*See* pp. 158-183, 359-363.)

[16] This step was taken first by Bar-Hillel (*See "Language and Information"*, esp. pp. 61 ff, 77, 101, 187).

[17] Rule (I - i) is adapted from Bochenski, *op. cit.*, p. 73. We must distinguish between *"A and B are not belonging to the same SC"* and *"A and B belong to different SCs"*; in conjunction with the premiss that A and B are well-formed expressions, the first implies the second, but not *vice versa.* A and B may belong both *to different SCs* and *to the same SC.* That would be the case e.g., if A is ascribed to two different SCs $\alpha$ and $\beta$, while B is ascribed to $\beta$ but not to $\alpha$. We see that it is possible that A and B as well as B and C belong to the same SC while A and C are not belonging to the same SC. Consequently, "belong to the same SC" is a *non-transitive relation.* It is however *reflexive* and *symmetical*, i.e., a similarity.

[18] Bar-Hillel was the first to extend Ajdukiewicz's system by the introduction of *"bidirectional"* SCs (as well as by the assignement of more than one SC to one and the same phrase). On the other hand, Bar-Hillel (after having followed Ajdukiewicz in admitting operators with more than one operands) actually allows

only one-place operators; i.e., he admits only operator SCs of the form $\lceil\alpha/\beta\rceil$ and $\lceil\beta\backslash\alpha\rceil$. This seems to be a vindication of the classical Aristotelian *subject-predicate* interpretation of propositions. I am not following Bar-Hillel in this respect, for the sake of preserving the relational interpretation of sentences like *'John loves Mary'* (even at the cost of enduring certain difficulties). *Cf.* Bar-Hillel, *op. cit.*, pp. 101, 191.

[19] *Cf.* Bar-Hillel, *op. cit.*, esp. pp. 100 f, 188.

[20] *See* Bar-Hillel, *op.cit.*, p. 61.

[21] *See* Bar-Hillel, *op. cit.*, pp. 61, 63, 77.

[22] "Nominals, verbals, adjectivals, etc... are syntactical categories" as Bar-Hillel put it, "they should not be confused with nouns, verbs, adjectives, etc., which are *morphological (paradigmatic) categories*". *See op. cit.*, p. 76.

[23] *'Man'* as a transitive verb belongs to $[n\backslash s/n]$; *'ball'* as an intransitive verb belongs to $\lceil n\backslash s\rceil$ and as transitive verb to $[n\backslash s/n]$.

[24] This example is taken from Bar-Hillel, *op. cit.*, p. 78-79.

[25] *Cf.* Bar-Hillel, *op. cit.*, esp. p. 187; Chomsky, op. cit., pp. 2, 48 ff; Gross et Lentin, "*Notions sur les Grammaires Formelles*", pp. 17, 109 ff.