THE JOURNAL OF SYMBOLIC LOGIC
Volume 18, Number 3, Sept. 1953

SYSTEMS OF SYNTACTIC ANALYSIS

NOAM CHOMSKY

1. Introduction.[1] During the past several decades, linguists have developed and applied widely techniques which enable them, to a considerable extent, to determine and state the structure of natural languages without semantic reference. It is of interest to inquire seriously into the formality of linguistic method and the adequacy of whatever part of it can be made purely formal, and to examine the possibilities of applying it, as has occasionally been suggested,[2] to a wider range of problems. In order to pursue these aims it is first necessary to reconstruct carefully the set of procedures by which the linguist derives the statements of a linguistic grammar from the behaviour of language users, distinguishing clearly between formal and experimental in such a way that grammatical notions, appearing as definienda in a constructional system, will be formally derivable for any language from a fixed sample of linguistic material upon which the primitives of the system are experimentally defined. The present paper will be an attempt to formalize a certain part[3] of the linguist's generalized syntax language.

From another point of view, this paper is an attempt to develop an adequate notion of syntactic category within an inscriptional nominalistic framework. The inscriptional approach seems natural for linguistics, particularly in view of the fact that an adequate extension of the results of this paper will have to deal with the problem of homonymity, i.e., with a statement of the conditions under which tokens of the same type must be assigned to different syntactic classes. It will appear below that the calculus of individuals can often supply quite simple solutions to constructional

Received October 18, 1952.

[1] Within linguistics, the source for these investigations is in the methods of structural analysis developed by Z. S. Harris; within philosophy and logic, it is in the work of N. Goodman on constructional systems and in the development of nominalistic syntax by Goodman and Quine. As general references, then, for this paper, see HARRIS, ***Methods in structural linguistics,*** Chicago, 1951, GOODMAN, ***The structure of appearance,*** Cambridge, 1951, and GOODMAN and QUINE, *Steps towards a constructive nominalism,* this JOURNAL, vol. 12 (1947), pp. 105–122. I am much indebted to Professors Harris, Goodman, and Quine, as well as to Y. Bar-Hillel, H. Hiż, and others, for many suggestions and criticisms.

[2] E.g., W. V. QUINE, *Notes on existence and necessity,* ***Journal of philosophy,*** vol. 40 (1943), pp. 120. Also, see Z. S. HARRIS, *Discourse analysis,* ***Language,*** vol. 28 (1952), pp. 1–30, for an investigation of the possibility of using methods of linguistics to determine the structure of a connected short text, thus, in a sense, setting up partial synonymity classes for it.

[3] The constructions of this paper are roughly coextensive with the procedures of chapters 15, 16, ***Methods.***

problems that seem on the surface to require a set-theoretic solution, thus removing the necessity for an involved hierarchy of types and increasing the overall workability of the system.

There are several ways in which we might approach the concept 'belong to the same syntactic category.' We might consider assigning elements to the same category only if they have all contexts in common (i.e., are mutually substitutable in all sentences), if they share some context, or if the ancestral of this latter relation holds between them.[4] All three suggestions are too restrictive for the general case which we wish to consider.[5] In particular, consider the following six-sentence text:

(1) 'ab', 'cb', 'de', 'fe', 'axd', 'cyf'.

To attain the purposes of this constructional attempt, we must be able to assign 'x' and 'y' to the same category. The general procedure which we wish to reconstruct is roughly as follows. If, in a given body of material, two elements occur in sentences which differ only in these elements (e.g., 'a' and 'c' occur in the context '..b', 'd' and 'f' in '..e', in (1)), then the two elements are assigned to the same class. But now two expressions differing term by term only in elements previously assigned to the same class (e.g., 'a..d' and 'c..f') are identified, thus shrinking the totality of contexts and allowing new elements (e.g., 'x' and 'y') to be put into the same class on the basis of occurrence in the same sentential context. When this process can be carried no further, considering expressions of any length and degree of discontinuity as elements, the resulting classes are the broadest syntactic categories for this text.

Before proceeding with the actual constructions, it should be made clear that the present system as given here is not adequate for the analysis of natural languages. Several crucial problems have been explicitly avoided in this treatment. One is the problem of homonymity touched on above. The second is the problem posed by those sentential contexts in which members of various syntactic classes can occur, e.g., in English, 'it was ....' The third concerns the predictive character of the grammar. A syntactic analysis will result in a system of rules stating the permitted sequences of the syntactic categories of the analyzed sample of the language, and thus generating the possible or grammatical sentences of the language. We can state this problem as one of determining the limitations on distribution which characterize various subsets of the broad syntactic categories, and of determining for any given language which subsets should be established.

---

[4] See Y. BAR-HILLEL, *On syntactic categories*, this JOURNAL, vol. 15 (1950), pp. 1–16, for a development of these notions.

[5] The third suggestion is actually equivalent to the system adopted here for the special case of languages in which each sentence contains exactly two elements (morphemes).

These problems lie beyond the scope of this paper,[6] which is concerned solely with the derivation, from reasonably limited samples,[7] of the broadest categories and the most general statement of the grammatical rules. These problems are related in various ways, and their solution seems to require auxiliary systems using richer means than those adopted here. Thus in actual linguistic procedure, problems of the second type, for example, are avoided by considerations involving the relative size of the logical product of the sets of contexts (distributions) of the elements involved.[8] The final decision as to the general adequacy of the systems proposed here will of course depend on the success or failure of such auxiliary systems.

2. Elementary notions of syntax. We therefore take as the general apparatus for these constructions quantification theory with variables ranging over inscriptions, that is, morphemes[9] and sums (scattered or continuous)[10] of morphemes. As extra-logical primitives we take the following.

'O', read 'overlaps.' This is the primitive notion of the calculus of individuals.[10] Two inscriptions overlap if there is some one inscription that is a part of both.

'EQL', read 'equally long.' Two inscriptions are equally long if they contain exactly the same number of morphemes (atoms).

[6] The first two in particular are problems of how to apply the primitives of these systems. Thus 'CON' must not be predicated of homonyms, and 'ENV' must not be predicated of contexts such as 'it was ...' (see below, § 2).

[7] Thus we do not wish to require in principle that the 'whole language' be available as data. It is, however, of interest to consider this situation as well. Thus, if there are large significant classes which are subdivided into classes whose distributions cluster separately (see footnote 8), but such that the subclasses have similar distributions in terms of other *classes*, then the methods to be adopted here permit the construction of the large class as a 'second-level' class.

[8] Thus we might require, for an expression to be admissible into the class of contexts, that the distributions of the elements occurring in its 'blank space' form a single cluster of sets. It is therefore necessary on the one hand to clarify the sense in which a set of sets can be said to be most efficiently divided into a set of clusters of sets, on the other, to investigate the actual statistics of distribution in natural languages. Precisely the same researches are necessary to resolve at least part of the homonym problem, considering homonyms as the elements whose distributions overlap two clusters of distributions. Cf. HARRIS, ***Methods,*** pp. 257ff.

[9] Actually, over morpheme occurrences. The linguist's morphemes are classes of conforming minimal meaning-bearing units, e.g., 'boy,' 'think,' 'of,' 'ing,' the plural 's', etc. Forms such as 'wife' and 'wive,' with selection predictable given the context (thus 'wive' occurs only before 's' plural, 'wife' only elsewhere), are called morpheme alternants and are considered to belong to the same morpheme. They are here considered to conform. See ***Methods,*** chap. 12, 13.

[10] For a discussion of the Calculus of Individuals (and the notions of 'sum', 'scattered individual', etc.) see H. S. LEONARD and N. GOODMAN, *The calculus of individuals and its uses*, this JOURNAL, vol. 5 (1940), pp. 45–55, and ***Structure,*** pp. 42–55.

'CON', read 'conform.' This is an equivalence relation holding between inscriptions of any length and shape (as long as they contain discontinuities at exactly the same places) which are, atom by atom, tokens of the same type, with the special considerations cited above in the case of homonyms and morpheme alternants.[9]

'PRE', read 'precedes.' This is a total ordering among atoms, a partial ordering among longer inscriptions (see A6, A7, below).

'ENV', read 'environment.' The long inscription (the text) to be analyzed is divided into environments such that if the sum of *a* and *b* (non-overlapping) is an environment, then *a* is a context of *b* and vice versa. Thus it might be useful to take sentences as the environment system for natural languages.

We begin by constructing certain elementary notions.[11]

*D*1. 'SEG*ab*' for '$(x)(\mathrm{O}xa \supset \mathrm{O}xb)$'.
('*a* is a segment of *b*.')

*D*2. '$=ab$' for 'SEG*ab*.SEG*ba*'.
('*a* is identical with *b*.')

*D*3. '$\cap bc$' for '$(\imath a)(x)(\mathrm{SEG}xa \equiv .\mathrm{SEG}xb.\mathrm{SEG}xc)$'.
('the product of *b* and *c*.')

*D*4. '$\cup bc$' for '$(\imath a)(x)(\mathrm{O}xa \equiv .\mathrm{O}xb \vee \mathrm{O}xc)$'.
('the sum of *b* and *c*.')

*D*5. 'ATM*a*' for '$(x)(\mathrm{SEG}xa \supset =xa)$'.
('*a* is an atom.')

*D*6. 'TI*ab*' for '$\mathrm{SEG}ab.(x)(y)(z)(\mathrm{SEG}xa.\mathrm{SEG}ya.\mathrm{SEG}zb.\mathrm{PRE}xz.$
$\mathrm{PRE}zy. \supset \mathrm{SEG}za)$'.
('*a* is a through inscription of *b*.')

Thus if we have the inscription 'pqrstuv' and if *b* is 'pq..s..v', then 'pq', 'pq..s', 'q..s..v', etc., are through inscriptions of *b*, but not 'p..s'. Thus if TI*ab*, *a* need not be continuous unless *b* is. But now we can readily define '*a* is continuous.'

*D*7. 'CI*a*' for '$(x)(\mathrm{SEG}ax \supset \mathrm{TI}ax)$'.
('*a* is a continuous inscription.')

*D*8. 'BEG*ab*' for '$\mathrm{TI}ab.\sim(\mathrm{E}z)(\mathrm{SEG}zb.\mathrm{PRE}za)$'.
('*a* is a beginning of *b*.')

*D*9. 'END*ab*' for '$\mathrm{TI}ab.\sim(\mathrm{E}z)(\mathrm{SEG}zb.\mathrm{PRE}az)$'.
('*a* is an ending of *b*.')

*D*10. 'XL*ab*' for 'BEG*ab*.ATM*a*'.
('*a* is to the extreme left of *b*.')

[11] *D*1–4 are, respectively, D2.042, D2.044, D2.045, and D2.047 of ***Structure,*** pp. 44–46.

*D*11. 'XR*ab*' for 'END*ab*.ATM*a*'.
('*a* is to the extreme right of *b*.')

*D*12. 'MCS*ab*' for 'CI*a*.SEG*ab*.(*x*)(SEG*xb*.CI∪*ax*. ⊃ SEG*xa*)'.
('*a* is a maximal continuous segment of *b*.')

*D*13. 'DC*ab*' for '∼O*ab*.(E*x*)(E*y*)(SEG*xb*.PRE*xa*.SEG*yb*.PRE*ay*.CI∪*a*∪*xy*)'.
('*a* is a discontinuity of *b*.')

Thus in the example for *D*6, 'r' and 'tu' are the discontinuities of *b*. We assign to every inscription a discontinuity index consisting of one atom from each discontinuity, and one additional atom to ensure that even continuous inscriptions have a discontinuity index, there being no null inscription. Two inscriptions will then be equally discontinuous if their discontinuity indices are of the same length.

*D*14. '$\mathrm{I_{dc}}$*ab*' for '(*x*)(SEG*xa*.ATM*x*. ≡ .(E*y*)(DC*yb*.XL*xy*) ∨ XL*xb*)'.
('*a* is the discontinuity index of *b*.')

*D*15. '$\mathrm{E_{dc}}$*ab*' for '(*x*)(*y*)($\mathrm{I_{dc}}$*xa*.$\mathrm{I_{dc}}$*yb*. ⊃ EQL*xy*)'.
('*a* and *b* are equally discontinuous.')

*D*16. 'K*ab*' for 'ENV∪*ab*.∼O*ab*'.
('*a* is a context of *b*.')

With the aid of these concepts we can state an axiom system.[12]

*A*1. O*ab* ≡ (E*x*)(*y*)(O*yx* ⊃ .O*ya*.O*yb*).

*A*2. EQL*ab* ≡ :ATM*a*.ATM*b*. ∨ (E*c*)(E*d*)(E*e*)(E*f*)(EQL*ce*.EQL*df*.∼O*cd*.∼O*ef*. = *a*∪*cd*. = *b*∪*ef*).

*A*3. ATM*a* ⊃. CON*aa*.(CON*ab* ⊃ CON*ba*).(CON*ab*.CON*bc*. ⊃ CON*ac*).

*A*4. CON*ab* ≡ .EQL*ab*.(*x*)(*y*)(*r*)(*s*)(BEG*xa*.XR*rx*.BEG*yb*.XR*sy*.EQL*xy*. ⊃. CON*rs*.$\mathrm{E_{dc}}$*xy*).

*A*5. PRE*ab*.PRE*bc*. ⊃ PRE*ac*.

*A*6. PRE*ab*.SEG*xa*.SEG*yb*. ⊃ PRE*xy*.

*A*7. ATM*a*.ATM*b*. ⊃ .PRE*ab* ∨ PRE*ba* ∨ =*ab*.

*A*8. =*ab*.ENV*a*. ⊃ ENV*b*.

*A*9. =*ab*.CON*ac*. ⊃ CON*bc*.

*A*10. (E*z*)(=*z*∪*xy*).

---

[12] *A*1 and *A*10 are, respectively, 2.41 and 2.45 of ***Structure,*** pp. 44–46. The essential idea of *A*13 is discussed in ***Structure*** on pp. 47–48. This axiom system is adequate only if we assume that no inscription contains infinitely many atoms, and then carry out proofs in the metalanguage, using induction on the number of atoms in an inscription. Alternatively, we could adjoin several axioms involving 'EQL' which would permit the derivation of all theorems in which no schematically defined terms appear within the system.

*A*11. (E*x*)(E*y*)(XL*xz*.XR*yz*).

*A*12. ATM*x* ⊃ (E*y*)(E*z*)(CON*xy*.K*zy*).

*A*13. (E*x*)(..*x*..) ⊃ (E*a*)(*y*)(ATM*y* ⊃ .SEG*ya* ≡ (E*z*)(SEG*yz*. ..*z*..)).

We can now derive such expected theorems as the substitution rule for identity, 'O*ab* ≡ (E*x*)(=*x*∩*ab*)', '∼PRE*aa*', ''EQL', 'CON', and 'E_{dc}' are equivalences,' etc.

We may now proceed to the analysis of the concepts 'comparable' as applied to sequences, and 'same position', as applied to terms of comparable sequences. It will be necessary here to consider sequences as being bracketed into terms in certain specific ways.[13] We will say that two sequences are comparable (with respect to given bracketings) if they contain the same number of bracketed terms and the same number of discontinuities, and if these discontinuities occur in the same places with reference to the bracketed terms. E.g., the following are similar with respect to the bracketing indicated by space, where '...' is a discontinuity.[14]

(2) ab cdef ... h ... ijk l m
nopq r ... stu ... v wx yz

Two bracketed terms are in the same position in comparable sequences if each is the *n*th term. Thus 'comparable' will be a four-place and 'same position' a six-place predicate.

*D*17. 'DV*ab*' for '(E*r*)(BEG*ra*.SEG*rb*.(*x*)(*y*)(MCS*yb*.XR*xy*. ⊃ SEG*xr*))'. ('*a* is a divisor (or bracketing) of *b*.')

Thus any divisor of the first sequence in (2) will include at least the atoms 'f', 'h', and 'm', plus any number (or 0) of the other atoms of the sequence, plus any number (or 0) of atoms which may follow 'm', but none which precede 'a'. The divisor is at least as long as the discontinuity index. In *D*22 it will be convenient to consider a given divisor as being the divisor for a sequence and for all of its beginnings ending in an atom of the divisor. This explains why in *D*17 we consider not *a* itself, but only some beginning of it (*r*) to be included in *b*.

---

[13] The non-atomic terms will in the interesting cases be what are called 'immediate constituents' in linguistic terminology. Thus such a linguistic form as 'that poor fellow on the corner missed his bus' might be analyzed into two immediate constituents, a noun phrase ('that .... corner') and a verb phrase ('missed his bus'), in which case it might be shown to be equivalent in the sense of the procedure to be adopted to a sentence consisting simply of a noun and a verb, e.g., 'he fell.' These phrases in turn can be analyzed into immediate constituents (e.g., 'that poor fellow' and 'on the corner'), etc., until the ultimate constituents (morphemes) are reached. For a detailed discussion of constituent analysis and its problems see R. S. Wells, *Immediate constituents*, ***Language,*** vol. 23 (1947), pp. 81–117, and ***Methods.***

[14] For the time being, we restrict ourselves to terms which do not cross over discontinuities. See however Systems III, IV, V, pp. 15–18.

A term of a sequence is a maximal part of the sequence occurring between two atoms of the divisor and containing the second of these on its extreme right. Thus in (2), with the given bracketings, each sequence has six terms. 'Term' is in many ways the analogue of 'ϵ' in this system.

*D*18. 'TRM*abc*' for 'DV*cb*.(E*x*)(XR*xa*.=*x*∩*ac*).(BEG*ab* ∨ (E*y*)(SEG*yc*. PRE*ya*.TI∪*yab*))'.

('*a* is a term of *b* with respect to the divisor *c*.')

It follows that terms are continuous inscriptions.[14]

We assign to an atom in a sequence (with respect to a given divisor) a term index in such a way that the term index contains *n* atoms just in case the given atom is in the *n*th term of the sequence.

*D*19. '$\mathrm{I}_{\mathrm{trm}}abcd$' for 'DV*dc*.ATM*b*.SEG*bc*.(*x*)(SEG*xa* ≡ :PRE*xb*. SEG*xd*.∨ =*xb*)'.

('*a* is the term index of the atom *b* in *c* bracketed by *d*.')

Thus in the first sequence of (2), the term index of 'j' with the given bracketing is the discontinuous inscription 'b...f...h...j', and the index of 'a' is 'a' itself.

Two atoms are preceded by the same number of terms if their term indices are of the same length.

*D*20. '$\mathrm{E}_{\mathrm{trm}}abcdef$' for '(E*x*)(E*y*)($\mathrm{I}_{\mathrm{trm}}xabc$.$\mathrm{I}_{\mathrm{trm}}ydef$.EQL*xy*)'.

('the atoms *a* (in the sequence *b* bracketed by *c*) and *d* (in the sequence *e* bracketed by *f*) each occur in the *n*th term, for some *n*, of these sequences bracketed in this way.')

We can define a predicate analogous to 'E_{dc}' as follows.

*D*21. '$\mathrm{E}_{\mathrm{dv}}abcd$' for '(*x*)(*y*)(XR*xa*.XR*yc*. ⊃ $\mathrm{E}_{\mathrm{trm}}xabycd$)'.

('the sequence *a* is divided into the same number of terms by the divisor *b* as is the sequence *c* by the divisor *d*.')

We can now define 'comparable' and 'same position' in the sense described above.

*D*22. 'CMP*abcd*' for '$\mathrm{E}_{\mathrm{dv}}abcd$.(*x*)(*y*)(BEG*xa*.BEG*yc*.$\mathrm{E}_{\mathrm{dv}}xbyd$. ⊃ $\mathrm{E}_{\mathrm{dc}}xy$)'.

('the sequence *a* with divisor *b* is comparable to *c* with divisor *d*.')

*D*23. 'SP*abcdef*' for 'TRM*abc*.TRM*def*.CMP*bcef*.(*x*)(*y*)(XR*xa*.XR*yd*. ⊃ $\mathrm{E}_{\mathrm{trm}}xbcyef$)'.

('*a* and *d* are in the same position in *b* (bracketed by *c*) and *e* (bracketed by *f*), respectively.')

We see that '$\mathrm{E}_{\mathrm{trm}}$', '$\mathrm{E}_{\mathrm{dv}}$', 'CMP', and 'SP' are symmetrical and transitive, and are reflexive if their places are significantly filled.

Before proceeding with the actual formulation of the procedure of syntactic analysis, it will be useful to provide the following auxiliary notions.

*D*24. '$\mathrm{EI}_{\mathrm{trm}}ab$' for '$\mathrm{DV}ba.(x)(\mathrm{TRM}xab \supset (\mathrm{E}y)(\mathrm{E}w)(\mathrm{CON}xw.\mathrm{K}yw))$'.
('*a* is bracketed by *b* into terms conforming to environment-included[15] sequences.')

*D*25. '$\mathrm{EI}a$' for '$(x)(\mathrm{XR}xa \supset \mathrm{EI}_{\mathrm{trm}}ax)$'.
('*a* conforms to an environment-included sequence.')

*D*26. $\ulcorner \mathrm{EQ}_4(R)abcd \urcorner$ for $\ulcorner \mathrm{CMP}abcd.(x)(y)(\mathrm{SP}xabycd \supset Rxy)\urcorner$, where '*R*' is a syntactic variable ranging over two-place predicates.
('the sequences *a* and *c*, as bracketed by *b* and *d*, respectively, are *R*-equivalent.')

*D*27. $\ulcorner \mathrm{EQ}_2(R)ab \urcorner$ for $\ulcorner (\mathrm{E}c)(\mathrm{E}d)(\mathrm{EQ}_4(R)acbd)\urcorner$.
('the sequences *a* and *b* are *R*-equivalent.')

3. A system of syntactic analysis. The general procedure, stated in § 1, towards which we have been aiming, can be constructed within our system as follows. We construct an indefinite series of similarity relations ('S_n')[16] such that the even similarities ('S_{2n}') hold between sequences, and the odd similarities between terms of these sequences. 'S_0' holds between sequences that conform term by term.

*D*28. 'S_0abcd' for '$\mathrm{EQ}_4(\mathrm{CON})abcd$'.

*D*29. 'S_0ab' for '$\mathrm{EQ}_2(\mathrm{CON})ab$'.

Thus 'S_0' is simply conformity.[17]

'S_1' holds between terms which occur in contexts related by 'S_0'. 'S_2' will then hold between sequences which are S_1-equivalent, 'S_3' between terms which occur in contexts related by 'S_2', etc. In the example (1) in § 1, 'a' and 'c' would be related by 'S_1' (as would 'd' and 'f'), 'a..d' and 'c..f' by 'S_2', and 'x' and 'y' by 'S_3'. We can give the general definition of 'S_{2n}' as follows.

---

[15] 'environment-included' will always be used in the sense of proper inclusion.

[16] In the definitions themselves, the variables '*m*', '*n*', etc., must be taken as syntactic variables ranging over numerals; elsewhere (including range specification) it is convenient to take them as numerical variables, ranging over numbers.

[17] It thus appears that 'CON' as explained and axiomatized above could have been defined from a simple conformity relation among atoms. The same is true of 'PRE'. This conformity relation could, in turn, be defined as the ancestral of a non-transitive matching relation, in a way analogous to that demonstrated in ***Structure,*** pp. 234–235.

(*Added November* 19, 1952.) These reductions would in fact increase the complexity of the basis in the sense of ***Structure,*** pp. 59–85, because the predicate formed (in calculating complexity) by compounding 'EQL' and 'CON' would have two segments rather than one under this revision, since 'CON' would now hold only of atoms. However, under a more recent formulation of the notion of simplicity (N. GOODMAN, *New notes on simplicity*, this JOURNAL, vol. 17 (1952), pp. 189–191) the two bases would be of equal simplicity.

*D*30. $\ulcorner S_{2n}abcd \urcorner$ for $\ulcorner EQ_4(S_{2n-1})abcd \urcorner$ $(n > 0)$.

*D*31. $\ulcorner S_{2n}ab \urcorner$ for $\ulcorner EQ_2(S_{2n-1})ab \urcorner$ $(n > 0)$.

The four-place predicate will be designated by ''$S_{2n}(4)$'', the two-place, by ''$S_{2n}(2)$''. The following will also be useful.

*D*32. $\ulcorner \Sigma_{2n}ab \urcorner$ for $\ulcorner (x)(y)(XRxa.XRyb. \supset S_{2n}axby) \urcorner$ $(n \geq 0)$.

It remains to define 'S_{2n+1}'. We will see that there are various ways of constructing this definition, and that the kind of language for which the system may be adequate is directly determined by the choice teken here. The basic notion behind the definiens for '$S_{2n+1}ab$' in the first system to be considered, henceforth system I, is

(3) $(Ea_1)\,..\,(Ea_r)(S_{2n}aa_1.Ka_1a_2.S_{2n}a_2a_3 \ldots S_{2n}a_{r-2}a_{r-1}.Ka_{r-1}a_r.S_{2n}a_rb)$.

To ensure transitivity this must be strengthened slightly, replacing 'S_{2n}' by 'Σ_{2n}' at each end. Hence as the actual definition in system I we take the following.

*D*33. $\ulcorner S_{2n+1}ab \urcorner$ for $\ulcorner (Ea_1) \ldots (Ea_r)(\Sigma_{2n}aa_1.Ka_1a_2.S_{2n}a_2a_3.Ka_3a_4.$
$S_{2n}a_4a_5 \ldots S_{2n}a_{r-2}a_{r-1}.Ka_{r-1}a_r.\Sigma_{2n}a_rb) \urcorner$,
where $n \geq 0$, r being any multiple of 4
(i.e., $r=4i$, $i \geq 1$).

Thus for each n, '$S_{2n+1}ab$' is introduced as an abbreviation which stands, simultaneously and ambiguously, for each of infinitely many expressions, one for each multiple of 4. For each n, the definiens for '$S_{2n+1}ab$' may be understood as an infinite disjunction of terms, each term of the disjunction being of the form given schematically on the right of *D*33, there being, for each integer i, a term with $4i$ quantified variables. See, however, end of § 3.

We can now define 'same syntactic category'.

*D*34. '$SSCab$' for $\ulcorner S_nab \urcorner$, for some $n \geq 0$.
('*a* and *b* are in the same syntactic category.')

Since '$S_{2n}abcd \supset S_{2n}ac$' follows directly, we see that any two elements related by any one of the sequence of similarities are in the same syntactic category. The following theorems give several pertinent characteristics of system I.

*T*1. $EIa \equiv .(Ey)(Ew)(CONaw.Kyw).CIa$.

*T*2. 'S_{2n+1}' is symmetrical and transitive.

*T*3. '$S_{2n}(4)$' is symmetrical and transitive.

*T*4. '$S_{2n}(2)$' is reflexive and symmetrical.

*T*5. $S_{2n}abcd \supset S_{2n+2}abcd$ $(n \geq 1)$.

*T*6. $S_0abcd.EI_{trm}ab. \supset S_2abcd$.

*T*7. $S_{2n}ab \supset S_{2n+2}ab$ $(n \geqq 0)$.

*T*8. $S_{2n+1}ab \supset S_{2n+3}ab$ $(n \geqq 0)$.

*T*9. $EIa.EIb. \supset .S_{2n}ab \supset S_{2n+1}ab$ $(n \geqq 0)$.

*T*10. $EIa \supset S_{2n+1}aa$.

*T*11. $EI_{trm}ab \supset S_{2n}abab$.

*T*12. $S_{2n+1}ab \supset .EIa.EIb$.

*T*13. 'SSC' is an equivalence taken over the restricted field of inscriptions satisfying 'EI'.

The distinguishing feature for each system of analysis will be the restricted field partitioned by 'SSC' as defined in that system. From *T*1 and *T*13 we see that the restricted field associated with system I is the set of continuous inscriptions conforming to environment-included inscriptions. As the syntactic categories for a given language we may take the equivalence classes of the restricted field augmented by the addition of such inscriptions as bear 'SSC' to some member of that class, but bear 'SSC' to no member of any other class. This notion of an 'extended category' allows of several interpretations.

Suppose that $a_1, \ldots, a_n$ are inscriptions satisfying 'EI', and that $A_1, \ldots, A_n$ are the corresponding equivalence classes. Thus

(4) $A_i = \hat{x}(EIx.SSCxa_i)$.

If we require merely that the syntactic categories be disjoint, we may define the syntactic categories $\bar{A}_1, \ldots, \bar{A}_n$ as

(5) $\bar{A}_i = \hat{x}((Et)(t \epsilon A_i.SSCxt).(y)(SSCxy.EIy. \supset y \epsilon A_i))$.

If we require further that no member of a syntactic category bear 'SSC' to any member of any other, then we may take them as

(6) $\bar{A}_i = A_i \cup \hat{x}((Et)(t \epsilon A_i.SSCxt).(y)(z)(SSCxy.SSCyz.EIz. \supset z \epsilon A_i))$.

In either case we can state in non-class terms the definition of 'same extended category' ('SEC'). Thus along the lines of (5) we have

(7) '$SECab$' for '$(Et)(Eu)(EIt.EIu.SSCat.SSCbu.(x)(SSCax \vee SSCbx.EIx. \supset SSCxt))$'.

From (7) we can prove that 'SEC' is transitive. We see further that two inscriptions can be in the same extended category though not related by 'SSC'. The considerations of the next section do not depend on a decision as to the preferability of (5) or (6), or some third formulation, since there we will be concerned with general applicability of a system to all languages of a given kind, and thus will limit our attention to the restrictied fields corresponding to these kinds of language.

It is of interest to note that although we can define same syntactic category' and even 'same extended category' within the system, we cannot give an adequate definition of the term 'syntactic category' itself. A syntactic category cannot be considered a mere sum of inscriptions in the same syntactic category, since a given inscription may contain as part of itself inscriptions belonging to various syntactic categories, even the category to which it itself belongs.[18] This may not be serious if 'syntactic rule' (a term in the metalanguage to these systems, since syntactic rules are composed of expressions denoting syntactic categories, or, non-univocally, their members) can be defined directly in terms of 'SEC'.[19]

From the fact that the last few definitions were presented only schematically, because of the appearance of numerical variables, it is clear that we have defined 'same syntactic category' in a nominalistic system only in the sense that these constructions may be viewed as a form of grammatical systems. For any given finite amount of textual material, we can construct a definition of 'same syntactic category' by giving a finite realization of the 'rules' for the construction of definitions laid down in this form of systems. For each n, the definiens of $D33$ will be a disjunction of a finite number of terms, and there will be only finitely many n's for which 'S_{2n+1}' need be defined.

4. Alternative systems. Adequacy. Suppose that we have a classification of languages on some structural basis, and a set of criteria of adequacy which must be met by the concept 'same syntactic category' as defined in a system of analysis. We will say that such a system is *applicable* to a given kind of language if the defined term meets the criteria of adequacy when applied to any language of this kind. Two systems are *equivalent for a given kind of language* if they are each applicable to this kind of language and if they yield exactly the same syntactic categories for the constituents[20] of the language when applied to any language of this kind. Two systems are *equivalent* if they are equivalent for all kinds of language to which either is applicable. System B is an *extension* of System A if it is equivalent to system A for all kinds of language to which system A is applicable, and is applicable to some kind of language to which system A is not applicable.

We can initiate a very limited investigation into these questions by making a simple classification of languages into kind 1, with only continuous constituents, and kind 2, with at least some discontinuous constituents, considering here only languages whose constituents conform to sequences

---

[18] As in the so-called *endocentric* constructions, e.g., 'poor John,' which belongs to the same category as 'John.' See L. BLOOMFIELD, ***Language,*** p. 194.

[19] It seems that this can be done by means of the devices developed by R. M. MARTIN and J. H. WOODGER, *Towards an inscriptional semantics*, this JOURNAL, vol. 16 (1951), pp. 191–203.

[20] We will call the immediate constituents, their immediate constituents, etc., down to ultimate constituents, simply the *constituents* of the language.

properly included in environments. As a criterion of adequacy we will here take the simple (and in general, obviously insufficient) condition that the syntactic categories be a partitioning of the constituents of the language. Thus system I is applicable to languages of kind 1 only. We can now construct systems which are extensions of system I, and systems applicable to both kinds of language although not extensions of system I.

Consider system II, constructed from system I by replacing *D*33 by (3), § 3, as *D*33^2.[21] Thus *D*33^2 differs from *D*33 only in that it has 'S_{2n}' in both places in which *D*33 has 'Σ_{2n}'. We can establish

*T*14. $\mathrm{EI}a.\mathrm{EI}b. \supset .\mathrm{SSC}ab \equiv \mathrm{SSC}^2ab.$

*T*15. 'SSC^2' is an equivalence over the restricted field of inscriptions, continuous or discontinuous, conforming to inscriptions properly included in environments.

Thus by relaxing this one restriction on 'S_{2n+1}' we derive an extension of system I. Although 'S^2_{2n+1}' is not generally transitive, we see that it is transitive in the interesting cases.

We may now attempt further and more interesting simplifications. Consider system III, based on *D*33^3.

D33^3. $\ulcorner \mathrm{S}^3_{2n+1}ab \urcorner$ for $\ulcorner(\mathrm{E}a_1)(\mathrm{E}a_2)(\mathrm{E}a_3)(\mathrm{E}a_4)(\Sigma^3_{2n}aa_1 . \mathrm{K}a_1a_2 . \mathrm{S}^3_{2n}a_2a_3 . \mathrm{K}a_3a_4 . \Sigma^3_{2n}a_4b) \quad (n \geq 0).$

This is just *D*33 with $i = 1$. Since 'S^3_{2n+1}' is not transitive, a basic problem will be to show that SSC^3ac, where

(8) $\Sigma^3_{2n}aa_1 . \mathrm{K}a_1a_2 . \mathrm{S}^3_{2n}a_2a_3 . \mathrm{K}a_3a_4 . \Sigma^3_{2n}a_4b$, and
$\Sigma^3_{2n}bb_1 . \mathrm{K}b_1b_2 . \mathrm{S}^3_{2n}b_2b_3 . \mathrm{K}b_3b_4 . \Sigma^3_{2n}b_4c$.

Basically, the line of proof will be as follows. We show that (i) for some m, $\mathrm{S}^3_{2m}a_4b_1$, $\Sigma^3_{2m}a_2a_3$, and $\Sigma^3_{2m}b_2b_3$. It follows that (ii) $S^3_{2m+1}a_2b_3$, by *D*33^3, and further, that (iii) $\mathrm{S}^3_{2m+2}a_2b_3$. But now we can show that (iv) $\Sigma^3_{2m+2}aa_1$ and $\Sigma^3_{2m+2}b_4c$, hence, with (iii) and '$\mathrm{K}a_1a_2 . \mathrm{K}b_3b_4$' from (8), we show, by *D*33^3, that (v) $\mathrm{S}^3_{2m+3}ac$.

But suppose that a_2 is discontinuous. We derive (iii) from (ii) by considering a_2 and b_3 as single terms (i.e., instead of (iii) we actually have proved the stronger '$\Sigma^3_{2m+2}a_2b_3$'). Thus we must permit discontinuous sequences to be terms in system III. This can be effected most simply by revising *D*17, dropping the requirement that a divisor must include the atom on the extreme right of each maximal continuous segment.

D17^3. 'DV^3ab' for '$(\mathrm{E}r)(\mathrm{BEG}ra . \mathrm{SEG}rb . (x)(\mathrm{XR}xb \supset \mathrm{SEG}xr))$'.

---

[21] The systems constructed in this section will keep the symbolism of system I (as well as the numbering of definitions and theorems), but with numerical superscripts, '2' for system II, etc. The symbols of system I appear without superscripts. Obviously, $\mathrm{K}ab \equiv \mathrm{K}^2ab$, $\mathrm{EI}a \equiv \mathrm{EI}^2a$, etc. Superscripts will ordinarily be dropped in such cases.

But again if a_2 is discontinuous, since 'S_{2n}' holds only between comparable (hence equally discontinuous) sequences, to assert (iii) we must revise *D*22 to permit a continuous and a discontinuous sequence to be comparable[3] in this special case.

$$D22^3.\ \text{`CMP}^3abcd\text{' for `}E^3_{dv}abcd.(x)(y)(\text{BEG}xa.\text{BEG}yc.E^3_{dv}xbyd.\supset E_{dc}xy).\vee.\text{XR}ba.\text{XR}dc\text{'}.$$

System III is then identical with system I except that *D*17, *D*22, and *D*33 are replaced, respectively, by $D17^3$, $D22^3$, and $D33^3$. We can now establish

$T1^3$. $\text{EI}^3a \equiv (\text{E}y)(\text{E}w)(\text{CON}aw.\text{K}yw)$.

T16. $\text{EI}^3a.\text{EI}^3b.\text{EI}^3c.\supset:S^3_nab.S^3_nbc.\supset \text{SSC}^3ac$.

*T*17. 'SSC^3' is an equivalence taken over the restricted field of inscriptions satisfying 'EI^3'.

Thus system III is applicable to both kinds of language. Before investigating the relation between systems II and III, we consider system IV, bearing the same relation to system III as did II to I, i.e., differing from III only in that $D33^3$ is replaced by $D33^4$, where $D33^4$ is just $D33^3$ with 'Σ_{2n}' replaced by 'S_{2n}' throughout. Systems III and IV are equivalent, thus

*T*18. $\text{EI}^3a.\text{EI}^3b.\supset.\text{SSC}^3ab \equiv \text{SSC}^4ab$.

Relating systems I and II with system III we have the following, as the strongest such theorems.

*T*19. $\text{EI}a.\text{EI}b.\supset.\text{SSC}ab \supset \text{SSC}^3ab$.

*T*20. $\text{EI}^3a.\text{EI}^3b.\supset.\text{SSC}^2ab \supset \text{SSC}^3ab$.

Thus system III is not an extension of system I. This is illustrated in the following four-sentence text, in which primed and unprimed terms conform.

(9) 'azb', 'cz′d', 'a′rb′x', 'c′sd′x′ '.

On the basis of the first two sentences we can assert that 'S^2_1' and 'S^3_1' hold between 'a .. b' and 'c .. d' (and trivially, between 'x' and 'x″). Hence 'S^3_2' holds between 'a′ .. b′x' (bracketed into 'a′ .. b″ and 'x') and 'c′ .. d′x″ (bracketed into 'c′ .. d″ and 'x″). Hence 'S^3_3' and 'SSC^3' hold between 'r' and 's'. But 'S^2_2' cannot hold between 'a′ .. b′x' and 'c′ .. d′x″ bracketed in this way, since discontinuous inscriptions cannot be terms. Hence neither 'SSC' nor 'SSC^2' holds between 'r' and 's'.

Of course we would naturally be inclined to say that a text like (9) belongs to a language with discontinuous constituents. This consideration suggests that it might be of interest to invert the procedure of this section and to attempt to determine the structural classification of languages formally in terms of the results given by application to them of various systems.

Thus the distinguishing characteristic of a language of kind 1, as opposed to kind 2, would seem to be that its continuous inscriptions are partitioned by system I in exactly the same way as by system III.

Although system III failed to be an extension of system I we can construct a system equivalent to system II even with the essential simplification of *D*33 which characterized system III. System V is the same as system III except that *D*30³ and *D*31³ are replaced as follows.

*D*30⁵. $\ulcorner S^5_{2n}abcd\urcorner$ for $\ulcorner EQ^5_4(S^5_{2n-1})abcd.XRba \vee (x)(y)(SP^5xabycd \supset . CIx.CIy)\urcorner$ $(n \geq 1)$.

*D*31⁵. $\ulcorner S^5_{2n}ab\urcorner$ for $\ulcorner(Ec)(Ed)(S^5_{2n}acbd)\urcorner$.

*T*21. $EI^5a.EI^5b. \supset .SSC^2ab \equiv SSC^5ab$.

This is the required theorem, since 'EI⁵' (or equivalently, 'EI³') is the condition that defines the restricted field of 'SSC²'.

Consider system VI, identical with system I except that *D*33 and *D*34 are replaced as follows.

*D*33⁶. $\ulcorner S^6_{2n+1}ab\urcorner$ for $\ulcorner(Ea_1) .. (Ea_r)(CONaa_1.Ka_1a_2.S^6_{2n}a_2a_3.Ka_3a_4. CONa_4a_5.Ka_5a_6.S^6_{2n}a_6a_7.Ka_7a_8.CONa_8a_9 \ldots CONa_{r-4}a_{r-3}.Ka_{r-3}a_{r-2}.S^6_{2n}a_{r-2}a_{r-1}.Ka_{r-1}a_r. CONa_rb)\urcorner$, where $n \geq 0$, $r=4i$, $i \geq 1$.

*D*34⁶. 'SSC⁶*ab*' for $\ulcorner S^6_{2n+1}ab\urcorner$, for some $n \geq 0$.

*T*22. $SSC^6ab \supset SSC^2ab$.

Since system VI partitions the inscriptions satisfying 'EI³', it gives a subpartitioning of that given by system II. That a significantly different notion of 'syntactic category' is involved here can be seen from the fact that *T*9⁶ does not hold. It was this that necessitated the change of *D*34.

Finally, consider system VII, differing from system I only in that *D*33 is replaced by *D*33⁷.

D33⁷. $\ulcorner S^7_{2n+1}ab\urcorner$ for $\ulcorner(Ea_1) \ldots (Ea_r)(CONaa_1.Ka_1a_2.S^7_{2n}a_2a_3.Ka_3a_4. S^7_{2n}a_4a_5 \ldots S^7_{2n}a_{r-2}a_{r-1}.Ka_{r-1}a_r.CONa_rb)\urcorner$, where $n \geq 0$, $r=4i$, $i \geq 1$.

*T*23. $EI^3a.EI^3b. \supset .SSC^2ab \equiv SSC^7ab$.

These systems fall into two major descriptive groups. Group I, with an infinitely long definition of 'S_{2n+1}' for each *n*, contains systems I, II, VI, and VII. Group II, with a finite definition for each *n*, contains systems III, IV, and V. On the basis of equivalence of systems, we have a cross-classification into system VI alone, and of the remainder, system I and its extensions II, V, and VII on the one hand, and systems III and IV on the other. Although the systems of group II sacrifice the simplicity of the associated theorems of group I, they have several redeeming features. Since for every pair of elements in a syntactic category we can determine the lowest *n* such that 'S_n' holds between them, the limitation, for each *n*, of

'S_{2n+1}' results in a much finer subdivision of the syntactic categories into nested subcategories of increasingly divergent distribution, as reflected in increasing n. Furthermore it is clear that because of the restriction to continuous terms, the extensions of system I, although applicable in our terms, can not give interesting results in general for languages of kind 2, as is clearly shown by the failure of system II in the above analysis of (9).

The preceding discussion probably does not exhaust the constructional possibilities within the limits of this general program, but until a more refined set of criteria of adequacy is given, and along with it, a refined analysis of kinds of language, it seems premature to enter into an exhaustive analysis of all possible systems of this general form. Furthermore, there are certain peculiar special features of these systems. Thus if CONaa', CONbb', etc., we see that from 'ax' and 'ya'' (where a = a', etc.), we can derive 'S_1xy', in fact, '$S_1x'y''$'; but we cannot do so from such pairs as 'abxc' and 'a'yb'c'', etc. It appears as well that even though system III is applicable to languages of kind 2, it cannot give interesting results for certain languages of this kind because of the limitation of terms of a sequence to through inscriptions. In both cases attempts at more general solutions lead to problems which are difficult to meet in the absence of clearer notions of adequacy.

SOCIETY OF FELLOWS, HARVARD UNIVERSITY