63-3-1

FOUR LECTURES ON ALGEBRAIC LINGUISTICS AND MACHINE TRANSLATION

A revised version of a series of lectures given in July, 1962 before a NATO Advanced Summer Institute on Automatic Translation of Languages in Venice, Italy.

by
Yehoshua Bar-Hillel
Professor of Logic and Philosophy
of Science
Hebrew University, Jerusalem, Israel

January, 1963

The work upon which these lectures are based was supported by the United States Office of Naval Research, Information Systems Branch under Contract N 62558-2214

FOUR LECTURES ON ALGEBRAIC LINGUISTICS AND MACHINE TRANSLATION

A revised version of a series of lectures given in July, 1962 before a NATO Advanced Summer Institute on Automatic Translation of Languages in Venice, Italy

by

Yehoshua Bar-Hillel

Professor of Logic and Philosophy of Science

Hebrew University, Jerusalem Israel

January, 1963

The work upon which these lectures are based was supported by the U. S. Office of Naval Research, Information Systems Branch, under Contract N 62558- 2214.

TABLE OF CONTENTS

LECTURE 1: THE ROLE OF GRAMMATICAL MODELS IN MACHINE TRANSLATION

Linguistics, as every other empirical science, is a complex mixture of theory and observation. The precise nature of this mixture is still not too well understood, and in this respect the difference between linguistics and, say, physics is probably at most one of degree. This lack of methodological insight has often led to futile disputes between linguists and other scientists dealing with language, such as psychologists, logicians, or communication theoreticians, as well as among linguists themselves.

Recently, however, considerable progress has been made in the understanding of the function of theory in linguistics, as a result of which theoretical linguistics has come into full-fledged existence. Interestingly enough, the present customary name for this new subdiscipline is rather mathematical linguistics. This is slightly unfortunate: though the adjective 'mathematical' is quite all right if 'mathematical' is understood in the sense of 'theory of formal systems', which is indeed one of its many legitimate senses, it is misleading inasmuch as it is still associated, at least among the non-specialists, including the bulk of the linguists themselves, with numbers and quantitative treatment. That subdiscipline of linguistics, however, which deals with numbers and statistics should better be called statistical linguistics and rather carefully be kept apart from mathematical linguistics qua theoretical linguistics. Should one prefer to regard 'mathematical linguistics' as a term for a genus of which statistical linguistics is a species, then the other species should perhaps be named algebraic linguistics.

After this terminological aside which, I think, was not superfluous, let us briefly sketch the background and development of algebraic linguistics. In the hands of such authors as Harris [1] and Hockett [2] in the United States, Hjelmslev [3] and Uldall [4] in Europe, structural linguistics became more and more conscious of the chasm between theory and observation, and linguistic theory deliberately got an algebraic look. At the same time, Carnap [5] and the Polish logicians, especially Ajdukiewicz [6], developed the logical syntax of language which was, however, too much

preoccupied with rules of deduction, and too little with rules of formation, to exert a great influence on current linguistics. Finally, Post [7] succeeded in formally assimilating rules of formation to rules of deduction, thereby paving the way of the application of the recently developed powerful theory of recursive functions, a branch of mathematical logic, to all ordinary languages viewed as combinatorial systems [8], while Curry [9] became more and more aware of the implications of combinatorial logic to theoretical linguistics. It is, though, perhaps not too surprising that the ideas of Post and Curry should be no better known to professional linguists than those of Carnap and Ajdukiewicz.

It seems that a major change in the peaceful but uninspiring co-existence of structural linguists and syntax-oriented logicians came along when the idea of mechanizing the determination of syntactic structure began to take hold of the imagination of various authors. Though this idea was originally but a natural outcome of the professional preoccupation of a handful of linguists and logicians, it made an almost sensational break-through in the early fifties when it became connected with, and a corner-stone of, automatic translation between natural languages. At one stroke, structural linguistics had become useful. Just as mathematical logic, regarded for years as the most abstract and abstruse scientific discipline, became overnight an essential tool for the designer and programmer of electronic digital computers, so structural linguistics, regarded for years as the most abstract and speculative branch of linguistics, is now considered by many a must for the designer of automatic translation routines. The impact of this development was at times revolutionary and dramatic. In Soviet Russia, for instance, structural linguistics had, before 1954, unfailingly been condemned as idealistic, bourgeois and formalistic. However, when the Russian government awakened from its dogmatic slumber to the tune of the Georgetown University demonstration of machine translation in January 1954, structural linguistics became within a few weeks a discipline of high prestige and priority. And just as mathematical logic has its special

offspring to deal with digital computers, i.e., the theory of autcmata, so structural linguistics has its special offspring to deal with mechanical structure determination, i.e., algebraic linguistics, also called, when this application is particularly stressed, computational linguistics or mechano-linguistics. As a final surprise, it has recently turned out that these two disciplines, automata theory and algebraic linguistics, exhibit extremely close relationships which at times amount to practical identity.

To complete this historical sketch: around 1954, Chomsky, influenced by, and in constant exchange of ideas with Harris, started his investigations into a new typology of linguistic structures. In a series of publications, of which the booklet Syntactic Structures [10] is the best known, but also the least technical, he defined and constantly refined a complex hierarchy of such structures, meant to serve as models for natural languages with varying degrees of adequacy. Though models for the treatment of linguistic structures were also developed by many other authors, Chomsky's publications exhibited a degree of rigor and testability which was unheard of before that in the linguistic literature and therefore quickly became for many a standard of comparison for other contributions.

I shall now turn to a presentation of the work of the Jerusalem group in linguistic model theory before I continue with the description and evaluation of some other contributions to this field.

In 1937, while working on a master's thesis on the logical antinomies, I came across Ajdukiewicz's work [6]. Fourteen years later, having become acquainted in the meantime with structural linguistics, and especially with the work of Harris [1], and instigated by my work at that time on machine translation, I realized the importance of Ajdukiewicz's approach for the mechanization of the determination of syntactic structure, and published an adaptation of Ajdukiewicz's ideas [11].

The basic heuristic concept behind the type of grammar proposed in this paper , and later further developed by Lambek [12, 13, 14], myself [15] and

others, is the following: the grammar was meant to be a recognition (identification or operational) grammar, i.e., a device by which the syntactic structure, and in particular the sentencehood, of a given string of elements of a given language could be determined. This determination had to be formal, i.e., dependent exclusively on the shape and order of the elements, and preferably effective, i.e., leading after a finite number of steps to the decision as to the structure, or structures, of the given string. This aim was to be achieved by assuming that each of the finitely many elements of the given natural language had finitely many syntactic functions, by developing a suitable notation for these syntactical functions (or categories, as we became used to calling them, in the tradition of Aristotle, Husserl, and Leśniewski), and by designing an algorithm operating on this notation.

More specifically, the assumption was investigated that natural languages have what is known to linguists as a contiguous immediate-constituent structure, i.e., that every sentence can be parsed, according to finitely many rules, 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. This parsing was not supposed to be necessarily unique. Syntactically ambiguous sentences allowed for two or more different parsings. Examples should not be necessary here.

The variation introduced by Ajdukiewicz into this conception of linguistic structure, well known in a crude form already to elementary school students, was to regard the combination of constituents into constitutes (or syntagmata) not a concatenation inter pares but rather as the result of the operation of one of the constituents (the governor, in some terminologies) upon the others (the governed or dependent units). The specific form which the approach took with Ajdukiewicz was to assign to each word (or other appropriate element) of a given natural language a finite number of fundamental and/or operator categories and to employ an extremely simple set of rules operating upon these categories, so-called "cancellation" rules.

Just for the sake of illustration, let me give here the definition of bidirectional categorial grammar, in a slight variation of the one presented in a recent publication of our group [16]. We define it as an ordered quintuple $\langle V, C, \Sigma, R, \mathcal{U} \rangle$, where V is a finite set of elements (the vocabulary), C is the closure of a finite set of fundamental categories, say $\Psi_1, \ldots, \Psi_n$, under the operations of right and left diagonalization (i.e., whenever α and β are categories, $[\alpha/\beta]$ and $[\alpha\backslash\beta]$ are categories), Σ is a distinguished category of C (the category of sentences), R is the set of the two cancellation rules $[\varphi_i/\varphi_j], \varphi_j \longrightarrow \varphi_i$, and $\varphi_i, [\varphi_i\backslash\varphi_j] \rightarrow \varphi_j$, and $\mathcal{U}$ is a function from V to finite sets of C (the assignment function).

We say that a category sequence α directly cancels to β, if β results from α by one application of one of the cancellation rules, and that α cancels to β, if β results from α by finitely many applications of these rules (more exactly, if there exist category sequences $\gamma_1, \gamma_2, \ldots, \gamma_n$ such that $\alpha = \gamma_1$, $\beta = \gamma_n$, and γ_i directly cancels to γ_{i+1}, for i=1, ..., n-1).

A string $x = A_1 \ldots A_k$ over V is defined to be a sentence if, and only if, at least one of the category sequences assigned to x by $\mathcal{U}$ cancels to Σ. The set of all sentences is then the language determined (or represented) by the given categorial grammar. A language representable by such a grammar is a categorial language.

In addition to bidirectional categorial grammars, we also dealt with unidirectional categorial grammars, employing either right or left diagonalization only for the formation of categories, and more specifically with what we called restricted categorial grammars, whose set of categories consists only of the (finitely many) fundamental categories Ψ_i, and the operator categories $[\Psi_i\backslash\Psi_j]$ and $[\Psi_i\backslash[\Psi_i\backslash\Psi_k]]$ (or, alternatively, $[\Psi_i/\Psi_j]$ and $[\Psi_i/[\Psi_j/\Psi_k]]$).

One of the results obtained by Gaifman in 1959 was that every language determinable by a bidirectional categorial grammar can also be determined by a unidirectional grammar and even by a restricted categorial grammar.

A heuristically (though not essentially) different approach to the

formalization of immediate - constituent grammars was taken by Chomsky, within the framework of his general typology. He looked upon a grammar as a device, or a system of rules, for generating (or recursively enumerating) the class of all sentences. In particular, a context-free phrase structure grammar, a CF grammar for short, may be defined, again in slight variation from Chomsky's original definition, as an ordered quadruple $< V, T, S, P >$, where V is the (total) vocabulary, T (the terminal vocabulary) is a subset of V, S (the initial symbol) is a distinguished element of V-T (the auxiliary vocabulary), and P is a finite set of production rules of the form $X \rightarrow x$, where $X \in V - T$ and x is a string over V.

We say that a string x directly generates y, if y results from x by one application of one of the production rules, and that x generates y, if y results from x by finitely many applications of these rules (more exactly, if there exist sequences of strings $z_1, z_2, \ldots, z_n$ such that $x=z_1$, $y=z_n$ and z_i directly generates z_{i+1}, for $i=1, \ldots, n-1$).

A string over T is defined to be a sentence if it is generated by S. The set of all sentences is the language determined (or represented) by the given CF grammar.

My conjecture that the classes of CF languages and bidirectional categorial languages are identical - in other words, that for each CF language there exists a weakly equivalent bidirectional categorial language and vice versa - was proved in 1959 by Gaifman [16], by a method that is too complex to be described here. He proved, as a matter of fact, slightly more, namely that for each CF grammar there exists a weakly equivalent restricted categorial grammar and vice versa. The equivalent representation can in all cases be effectively obtained from the original representation.

This equivalence proof was preceded by another in which it was shown that the notion of a finite state grammar, FS grammar for short, occupying the lowest position in Chomsky's hierarchy of generation grammars, was equivalent to that of a finite automaton, in the sense of Rabin and Scott [17], which can

be viewed as another kind of recognition grammar. The proof itself was rather straightforward and almost trivial, relying mainly on the equivalence of deterministic and non-deterministic finite automata, shown by Rabin and Scott. It has been adequately described in a recently published paper [18].

Chomsky had already shown that the FS languages formed a proper subclass of the CF languages. We have recently been able to prove [19] that the problem whether a CF language is also representable by a FS grammar - a problem which has considerable linguistic importance - is recursively unsolvable. The method used was reduction to Post's correspondence problem, a famous problem in mathematical logic which was shown by Post [20] to be recursively unsolvable.

Among other results recently obtained, let me only mention the following: whereas FS languages are, in view of the equivalence of FS grammars to finite automata and well-known results of Kleene [21] and others, closed under various Boolean and other operations, CF languages whose vocabulary contains at least two symbols are not closed under complementation and intersection, though closed under various other operations. The union of two CF languages is again a CF language, and a representation can be effectively constructed from the given representation. The intersection of a CF language and a FS language is a CF language.

Undecidable are such problems as the equivalence problem between two CF grammars, the inclusion problem of languages represented by CF grammars, the problem of disjointedness of such languages, etc. In this connection, interesting relationships have been shown to exist between CF grammars and two-tape finite automata, as defined and treated by Rabin and Scott, for which the disjointedness problem of the sets of acceptable tapes is similarly unsolvable.

A particular proper subset of the CF languages, apparently of greater importance for the treatment of programming languages, such as ALGOL, than for natural languages, is the set of so-called sequential languages, studied in particular by Ginsburg [22, 23] and Shamir [24]. I have no time for more than just this remark.

In a somewhat different approach, closely related to the classical notions of government and syntagmata, the notions of dependency grammars and projective grammars have been developed by Hays [25], Lecerf [26], and others, including some Russian authors, utilizing ideas most fully presented in Tesnière's posthumous book [27], and are thought to be of particular importance for machine translation. However, it has not been too difficult to guess, and has indeed been rigorously proven by Gaifman [28], that these grammars, which are being discussed in other lectures presented in this Institute, are equivalent to CF grammars in a certain sense, which is somewhat stronger than the one used above, but that this is not necessarily so with regard to what might be called natural strong equivalence. More precisely, whereas for every dependency grammar there exists, and can be effectively constructed, a CF grammar naturally and strongly equivalent to it, this is not necessarily the case in the opposite direction, not if the CF grammar is of infinite degree. Let me add that the dependency grammars are very closely related to a type of categorial grammars which I discussed in earlier publications [11] but later on replaced by grammars of a seemingly simpler structure. In the original categorial grammars, I did consider categories of the form $\beta_m \ldots \beta_2 \beta_1 \backslash \alpha / \gamma_1 \gamma_2 \ldots \gamma_n$, with α, β_i, and γ_j being either fundamental or operator categories themselves, with a corresponding cancellation rule. It should be rather obvious how to transform a dependency grammar into a categorial grammar of this particular type. These grammars are equivalent to grammars in which all categories have the form $\beta \backslash \alpha / \gamma$ where α, β, and γ are fundamental categories and where β and γ may be empty (in which case the corresponding diagonal will be omitted, too, from the symbol). Finally, in view of Gaifman's theorem mentioned above, these grammars in their turn are equivalent to grammars all of whose categories are of the form $\alpha / \beta \gamma$ (or $\gamma \beta \backslash \alpha$), with the same conditions. I think that these remarks (strongly connected with considerations of combinatory logic [9]) should definitely settle the question of the exact formal status of the dependency grammars and their like. One side result is that dependency grammars are weakly reducible

to binary dependency grammars, i.e., grammars in which each unit governs at most two other units. This result, I presume, is not particularly surprising, especially if we remember that the equivalence proven will in general not be a natural one.

Still another class of grammars, sometimes [29] called push-down store grammars and originating, though not in a very precise form, with Yngve [30, 31], has recently been shown by Chomsky to be once more equivalent to CF grammars, again to nobody's particular surprise. Since push-down stores are regarded by many workers in the fields of MT and programming languages as particularly useful devices for the mechanical determination of syntactic structure of sentences belonging to natural and programming languages, respectively, this result should be helpful in clarifying the exact scope of those schemes of syntactic analysis which are based on these devices.

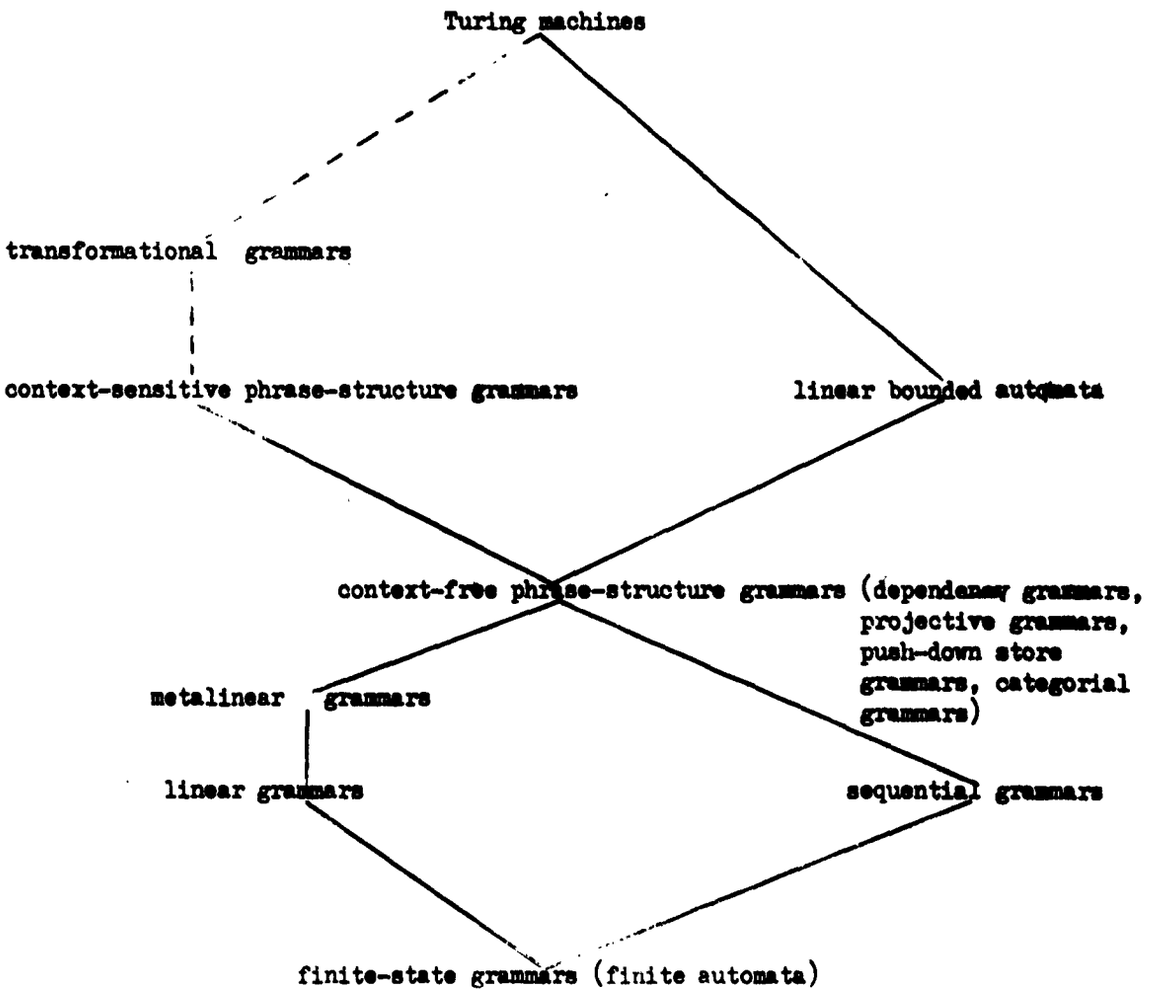
Of theoretically greater importance is the fact that push-down store grammars form a proper sub-set of linear bounded automata, one of the many classes of automata lying between Turing machines and finite automata which have recently been investigated by many authors, due to the fact that Turing machines are too idealized to be of much direct applicability, whereas finite automata are too restricted for this purpose.

The investigation of these automata, initiated by Myhill [32], is, however, still in its infancy, similar to that of many other classes of automata reported by McNaughton in his excellent review [33]. Still more in the dark is the linguistic relevance of all these models though, judging from admittedly limited experience, almost every single one of them will sooner or later be shown to have such relevance.

To wind up this discussion, let me only mention that during the last few years various classes of grammars whose potency is intermediate between FS and CF grammars have been investigated. These intermediate grammars will probably turn out to be of greater importance for the study of grammars of programming and other artificial formalized languages than for natural languages. In

addition to the sequential grammars mentioned before, let me now mention the linear and metalinear grammars studied by Chomsky.

It might be useful to present, at this stage, a picture of the various grammars discussed in the present lecture, together with the two important classes of transformational and context-sensitive phrase structure grammars (which I could not discuss, for lack of time) in the form of a directed graph based on the (partial) ordering relation Determine-a-more-extensive-class-of-languages-than (the staggered lines indicating that the exact relationship has not yet been fully determined):



The last two questions I would now like to discuss are the following: (1) In view of the fact that so many models of linguistic structure have turned out to be (weakly) equivalent, how do they compare from the point of view of pedagogy and MT-directed application? (2) What is the degree of adequacy with which natural languages can be described by CF grammars and their equivalents?

As to the first question, I am afraid that not much can be said at this stage. I am not aware of any experiments made as yet to determine the pedagogical status of the various equivalent grammars. Some programmatic statements have been made on occasion, but I would not want to attribute much weight to them. I myself, for instance, have a feeling that the governor-dependent terminology of the dependencyand projective grammars has an unfortunate, and intrinsically of course unwarranted, side-effect of strengthening dogmatic approaches to the decision of what governs what. The operator-operand terminology of the categorial grammars seems to be emotionally less loaded, but again, these are surely minor issues. Altogether, I would advocate the performance of pedagogical experiments in which the same miniature language would be taught with the help of various equivalent grammars. I do not foresee any particular complications for such projects.

Turning now to the second question which has been much discussed during the last few years, often with great fervor, the situation should be reasonably clear. FS grammars are definitely inadequate for describing any natural language, unless this last term is mutilated, for what must be regarded as arbitrary and ad hoc reasons. I am sorry that Yngve's otherwise extremely useful recent contributions did becloud this issue. As to CF grammars, the situation is more complex and more interesting. It is almost, but not quite, certain that such grammars, too, are inadequate in principle, for reasons which I shall not repeat here, since they have been stated many times in the recent literature and been authoritatively restated by Chomsky [28]. But of even greated importance, particularly for applications, such as MT, is the fact that such grammars seem definitely to be inadequate in practice, in the sense that the number and

complexity of grammatical rules of this type, in order to achieve a tolerable, if not perfect, degree of adequacy, will have to be so immense as to defeat the practical purpose of establishing these rules. Transformational grammars seem to have a much better chance of being both adequate and practical, though this point is still far from being settled. In view of this fact, which does not appear to have been seriously challenged by most workers on MT, it is surprising to see that most, if not all, current programs of automatic syntactic analysis are based on impractical grammars. In some groups, where the impracticability and/or inadequacy has received serious attention, attempts are being made at present to classify the "recalcitrant" phenomena and to find ad hoc remedies for them. You will not be surprised if I say that I take a rather dim view of these attempts. But this already leads to issues which I intend to discuss in subsequent lectures.

SECOND LECTURE: SYNTACTIC COMPLEXITY

Extremely little is known about syntactic complexity, though this notion has come up in many discussions of style, readability, and, more recently, of mechanisation of syntactic analysis. Its explication has been universally regarded as a matter of great difficulty, this probably being the reason why it has also been, to my knowledge, universally shunned. When such authors as Flesch [34] developed their readability measures, they could not help facing the problem but, unable to cope with it, replaced syntactic complexity in their formulas by length, whose measure poses incomparably fewer problems, while still standing in some high statistical correlation with the elusive syntactic complexity.

Very often one hears, or reads, of an author, a professional group, of even a whole linguistic community being accused of expressing themselves with greater syntactic complexity than necessary. Such slogans as "What can be said at all, can be said simply and clearly in any civilized language, or in a suitable system of symbols.", formulated by the British philosopher C. D. Broad in elaboration of a well-known dictum by Wittgenstein, were used by philosophers of certain schools to criticise philosophers of other schools, and have gained particular respectability in this context. On a less exalted level, most people interested in information processing and, in particular, in the condensation of information, preferably by machine, seem to be convinced that most, if not all, of what is ordinarily said could be said not only in syntactically <u>simpler</u> sentences but in syntactically <u>simple</u> sentences, the analysis of which would be a pleasure for a machine. Often, informationlossless transformation into syntactically simple sentences is regarded as a helpful, perhaps even necessary step prior to further processing. In the context of machine transation, Harris, e.g., once expressed the hunch that mechanical translation of kernel sentences, which would presumably rank lowest on any scale of syntactic complexity, should be a simpler affair than translation of any old sentences.

It is my conviction that the topic of syntactic complexity is, beyond certain very narrow limits of a vaguely felt concensus, ridden with bias, prejudice and fallacies to such a degree as to make almost everything that has been said on it completely worthless. In particular, I think that the "Wittgensteinian" slogan mentioned above is misleading in the extreme. I tend to believe that its attractiveness is due to its being understood not as a statement of fact but rather as a kind of general and vague advice to say whatever one wants to say as simply and clearly "as possible," something to which one could hardly object, though, as we shall see, even in this interpretation it is not unequivocally good advice, when simplicity is understood as syntactic simplicity, since the price to be paid for reducing syntactic complexity, even when it is "possible," may well turn out to be too high.

So far, I have been using "syntactic complexity" in its pretheoretical and unanalyzed vague sense. It is time to become more systematic.

One should not be surprised that the explication of syntactic complexity to which we shall presently turn will reveal that the pretheoretical term is highly equivocal, though one might well be surprised to learn <u>how</u> equivocal it is.

When I said in the opening phrase that "extremely little is known about syntactic complexity," I intended the modifier "extremely little" to be understood literally and not as a polite version of "nothing." Such terms as "nesting," "discontinuous constituents," "self-embedding" and "syntactic depth" are being used in increasing frequency by linguists in general and - perhaps unfortunately so - by applied linguists in particular, especially when programming for machine analysis is discussed. But not until very recently have these notions been provided with a reasonably rigid formal definition which alone makes possible their responsible discussion. The most recent and most elaborate discussion that has come to my attention is that by Chomsky and Miller [35]. They discuss there various explicata for "syntactic complexity," with

varying degrees of tentativeness, as befits such a first attempt, and I shall make much use of this treatment in what follows.

Let me first discard one notion which, as already mentioned, has a certain *prima facie* appeal to serve as a possible explicatum for syntactic complexity, namely *length*, measured, say, by the number of words in the sentence (or in whatever other construction is under investigation). Though, as said before, it is obvious that there should exist a fairly high statistical correlation between syntactic complexity and length, it should be equally obvious that length is entirely inadequate to serve as an explicatum for syntactic complexity. Take as many sentences as you wish of the form "... is ——" (such as "John is hungry.", "Paul is thirsty.", etc.) whose intuitive degree of syntactic complexity is close, if not equal, to the lowest one possible, join them by repeated occurrences of "and" (a procedure resulting in something like "John is hungry and Paul is thirsty and Mary is sleepy and..."), and you will get sentences of any length you wish whose intuitive degree of syntactic complexity should still be close to the minimum. True enough, a sentence of this form, containing 50 clauses of the type mentioned, always with different proper names in the first position and different adjectives in the third position would be difficult to remember exactly. Therefore such a sentence will be "complex," in one of the many senses of this word, but surely not syntactically so. No normal English-speaking person will have the slightest difficulty in telling the exact syntactic form, up to a parameter, of the resulting sentence, and there will be no increase in this difficulty even if the number of clauses will be 100, 1000, or any number you wish. In one very important sense of "understanding," the increased length of sentences of this type will *not* increase the difficulty of understanding them. And the sense in question is, of course, precisely that of grasping the syntactic structure.

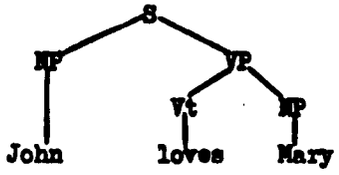
The next remark, prior to presenting some of the more interesting

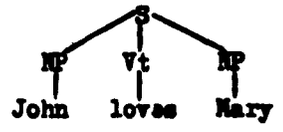
explicata, refers to a fact which I want very much to call to your careful attention. I hope it will not be as surprising to you as it was to me, the first time I hit upon it. For a time, I thought that the only relativization needed for explicating syntactic complexity would be the trivial one to a given language. (Logicians, and some linguists, know plenty of examples where the "same" sentence may belong to entirely different languages; in that case, nobody would be surprised to learn that it also has - or rather that they also have - different degrees of syntactic complexity, relative to their respective languages.) What did shock me, however, though only for a moment until I realized that it could not be otherwise, was that degree of complexity must also be explicated as being relative to a grammar, that the same sentence of the same language may have one degree of complexity when analyzed from the point of view of one grammar and a different one when analyzed from the point of view of another grammar, and that, of two different sentences, one may have a higher degree of complexity than the other relative to one grammar, but a lower degree relative to another grammar.

This doubtless being the case, may I be allowed a certain amount of speculation for a minute? It is a simple and well-known fact that the same sentence will sometimes be better understood by person A than by B, though they have about the same IQ, about the same background knowledge, and though they read or hear it with about equal attention, as far as one can make out. Could it be that they are (subconsciously, of course) analyzing this same sentence according to different grammars, relative to which this sentence has different degrees of syntactic complexity? Could it be that part of the improvement in understanding obtained through training and familiarization is due to the trainee's learning to employ another grammar (whose difference from the one he was accustomed to employ before might be only minimal, so that the acquisition of this new grammar might not have been too difficult, perhaps)? Could it be that many, if not all, of us work with more than one grammar simultaneously, switching from the one to the other

when the employment of the one runs us into trouble, e.g., when according to one grammar the degree of complexity of a given sentence is greater than one can stand? More about this later. Attractive as these speculations are, let me stress that at this moment I don't know of any way of putting them to a direct empirical test. But I wish someone would think up such a way. Let me also add that he who does not like this picture of different grammars for the same language lying peacefully side by side somewhere in our brain, may look upon the situation as one system of grammatical rules (the set-theoretical union of the two sets discussed so far) being stored in the brain, and allowing the same sentence to be analyzed and understood in two different ways with two different degrees of complexity, with a control element deciding which rules to apply in a given case and allowing the switch to other rules when trouble strikes. That there are syntactically ambiguous sentences has, of course, always been well known, but I am speaking at the moment about a particular kind of syntactic ambiguity, one that has no semantic ambiguities in its wake, but where the difference in the analysis still creates a difference in comprehensibility. At this point it is probably worthwhile to present an extremely simple example. The English sentence, "John loves Mary.", can be analyzed (and has been analyzed) in two different ways, each of which will be expressed here in two different but equivalent notations which have been simplified for our present purposes:

$(_S(_{NP}\text{John})(_{VP}(_{Vt}\text{loves})(_{NP}\text{Mary})))$ $\qquad$ $(_S(_{NP}\text{John})(_{Vt}\text{loves})(_{NP}\text{Mary}))$





These analyses correspond to the following two "grammars," G_1 and G_2:

G_1: S → NP + VP
VP → Vt + NP
NP → John, Mary
Vt → loves

G_2: S → NP + Vt + NP
NP → John, Mary
Vt → loves

or, if you prefer, they both correspond to the grammar G_3, which is the set-theoretical union of G_1 and G_2, and consists therefore of just the rules of G_1 plus the first rule of G_2. (Both G_1 and G_2 are of course CF- grammars; G_1 is binary, but G_2, and therefore also G_3, is not.)

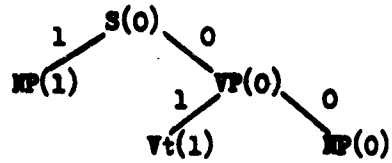
Though the difference in structure assigned to this sentence by the two analyses is palpable, it is less clear whether this difference implies a difference in the intuitive degree of syntactic complexity, and if so, according to which analysis the sentence is more complex. As a matter of fact, good reasons can be given for both views: in the first analysis, more rules are applied but each rule has a particularly simple form; in the second analysis fewer rules are applied, but one of them has a more complicated form. This situation seems to indicate that we have more than one explicandum before us, more than one notion which, in the pre-theoretical stage, is entitled to be called "syntactic complexity."

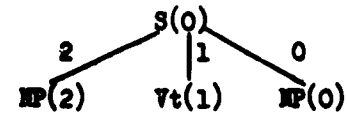
There are still more aspects to the intuitive uses of "syntactic complexity," but perhaps it is time to turn directly to the explicata which, hopefully, will take care of at least some of these aspects.

To follow Chomsky once again [35] rather closely, we might introduce the terms "depth of postponed symbols" and "node/terminal-node ratio" to denote the following two relevant measures: the first for Yngve's well-known depth-measure, which, I trust, will again be explained in his lectures at this Institute, the second for a new concept which has not yet been discussed in the literature. Both measures refer to the tree representing the sentence and are therefore applicable only to such grammars which assign tree structure to each sentence generated by them.

If we assign, in the Yngve fashion, numbers to the nodes and branches (with the branches leading to the terminal symbols left out), we see that the greatest number assigned to any of the nodes of the left tree is 1, so that its depth of postponed symbols is also 1, whereas the corresponding number for the second tree is 2. On the other hand, the total number of nodes of the first tree is 5, the number of its terminal nodes is 3, so

that its node/terminal-node ratio is 5/3, whereas the corresponding numbers for the second tree are 4, 3, and 4/3 respectively.



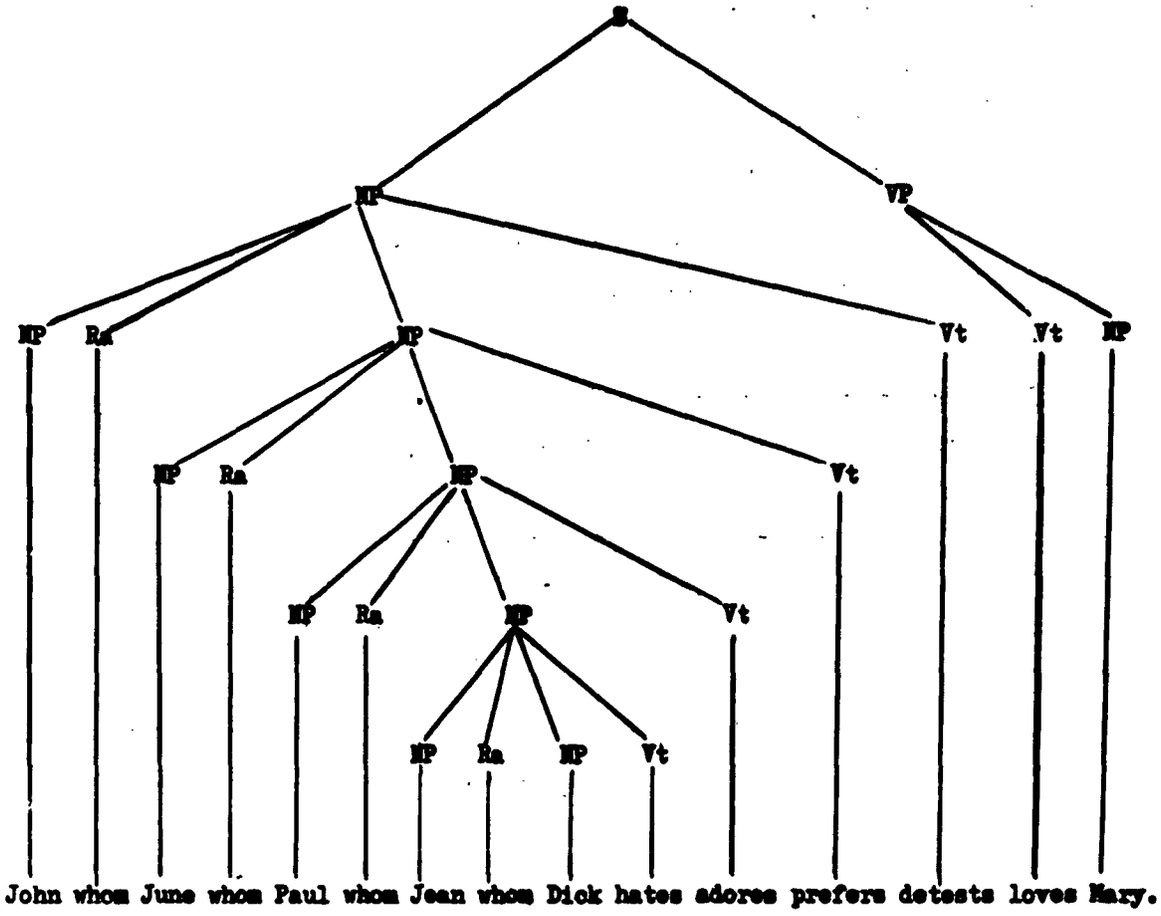


Each node number (in parentheses) is equal to the sum of the number assigned to the branch leading to this node and the number of the node from which the branch comes.

There are at least three more notions that are entitled to be considered as explicata for other aspects of syntactic complexity. The one that has been most studied is the degree of nesting. The reasons for the attention given to it are that it has been known for a long time that a highly nested sentence causes difficulties in comprehension and, more recently, that it creates troubles for mechanical syntactic analysis. One rough explication of this notion (there are others) might run as follows, again relative to tree grammars: The degree of nesting of a labeled tree is the largest integer m, such that there exists in this tree a path through m+1 nodes $N_0, N_1, \ldots, N_m$, with the same or different labels, where each N_i $(i \geq 1)$ is an inner node in the subtree rooted in N_{i-1}. The same degree of nesting is also assigned to the terminal expression as analyzed by this tree.

A special case of nesting is self-embedding, to whose importance Chomsky has called attention. In order to define the degree of self-embedding of a labeled tree, one has only to change in the above definition of degree of nesting the phrase "with the same or different labels" by the phrase "each with the same label." (Other definitions are again possible.)

To present one more stock example, the following tree has a degree of nesting (equal, in this particular case, to its degree of self-embedding) of 4. (Its depth, incidentally, is 7 and its node/terminal-node ratio is 21/15 = 7/5.)


S
NP
VP
NP
Ra
NP
Vt
Vt
NP
NP
Ra
NP
Vt
NP
Ra
NP
Vt
NP
Ra
NP
Vt
John whom June whom Paul whom Jean whom Dick hates adores prefers detests loves Mary.

Though this tree could have been derived from a grammar G_4 differing from G_3 only by containing the additional rules

$$NP \rightarrow NP + Ra + NP + Vt$$
$$Ra \rightarrow whom$$

there are very good reasons why sentences of the type

John whom Ann hates loves Mary

and their ramifications should, in the framework of the whole English language, not be regarded as being produced by a CF- grammar containing G_4 as a proper part, but rather by a transformational grammar built upon a CF- grammar of English containing, in addition, a transformational rule, which I shall not specify here, allowing the derivation of

$$NP_1 + Ra + NP_3 + Vt + Vt + NP_2$$

from

$$NP_1 + Vt + NP_2$$

and

$$NP_3 + Vt + NP_1 .$$

(There is no need to stress that all this is only a very rough approximation to the incomparably more refined treatment which a full-fledged transformational grammar of English would require. The transformational rule, for instance, should refer to the trees representing the strings under discussion rather than to the strings themselves.) It is worthwhile noticing that the node/terminal-node ratio (7/5) of the resulting tree is smaller than the ratios (5/3) of the underlying trees.

The fifth aspect of syntactic complexity is, then, <u>transformational history</u>. I am, of course, not using the term "measure" now, because it is very doubtful whether measures can be usefully assigned to this concept. So far, no attempt in this direction has been made. I shall, therefor, say no more about this notion here.

It is not particularly difficult to develop these five notions, and many more could be thought of. The decisive questions are twofold: What are the exact formal properties of the various notions and perhaps even

more important, what is their psychological reality, to use a term of Sapir's? In general, one would tend to require that if one sentence is syntactically more complex than another, then, *ceteris paribus*, it should, perhaps only on the average, create more difficulties in its comprehension. What can we say on this point?

Well, very little, and nothing so far under controlled experimental conditions. Highly nested constructions just don't occur at all in normal speech and very rarely in writing, with the notable exception of logical or mathematical formulas. Their syntactic structure can be grasped only by using extraordinary means such as going over them more than once and using special marks for pairing off expressions that belong together but between which other expressions have been nested. As formula such as

$$[[p \supset [q \supset [[r \supset [s \supset t]] \supset u]]] \supset v]$$

is certainly not a very complex one among the formulas of the propositional calculus, as they go, but testing its well-formedness would either require some artificial aids, such as the use of a pencil for marking off paired brackets, or the acquisition of a special algorithm based upon a particular counting procedure, or else just an extraordinary (and unanalysed) effort and concentration. It is doubtful whether any effort, without external aids, would suffice to determine that the "literal" English rendition of the formula as

If if p then if q then if if r then if s then t then u then v

is well-formed, when one listens to such a sentence without prior warning.

It is interesting that in order to explain our difficulties in either uttering or grasping the structure of such sentences we need assume nothing more than that we are finite automata with a finite number of internal states. For Chomsky [36], in effect, has shown that when the number of these states is some number n, then, relative to a given grammar G, there exists a number m (depending on n) such that this device will not be able to correctly analyze the syntactic structure of all sentences whose degree of nesting is greater than or equal to m. (As a matter of fact, Chomsky showed this for degree of self-embedding rather than for nesting, but the

proof can be trivially extended to this case.)

On the incomparably stronger assumption that natural languages (such as English) can be adequately determined by tree grammars, that human speakers of such a language have at least one such tree grammar stored in their permanent memory, that they utter the sentences of these languages by going through (one of) their tree(s) "from top to bottom and from left to right," that all storage required for this process is done in an immediate memory of the push-down store form containing, say, n cells, we arrive at the conclusion that only sentences whose depth of postponed symbols is no higher than n can be uttered by such speakers.

Now, though Yngve continues to believe that there exists good evidence for the soundness of these assumptions, Chomsky has on various occasions [37,38] expressed his doubts as to this evaluation of the evidence. He believes that most of the positive evidence invoked by Yngve can already be explained on the basis of the weaker assumption mentioned above, whereas he mentions the existence of other evidence which tends to refute Yngve's stronger assumptions though not his own weak one. I have no time to go further into this controversy. Let me only state that Chomsky's arguments seem to me to be the more conclusive ones. This, of course, by no means diminishes the credit due to Yngve for having been the first to have raised certain types of questions that were never asked before, and to have ventured to provide for them interesting answers, though they may well turn out to be the wrong ones.

It is time now to say at least a few words on the "Wittgensteinian Thesis." In one sense, this thesis is of course perfectly true: After all, all of us do manage to say most of what we have to say in sentences of a low degree of nesting and, if really necessary, could rephrase even those things for the expression of which we do use highly nested strings, such as occur in many mathematical formulas, in syntactically less complex ways, which will be presently investigated. But in this sense, the thesis is no more than a rather uninteresting truism. What Wittgenstein, Broad and the innumerably many other people who invoked this slogan doubtless had in mind

was that most, if not all, of the things that are expressed (usually, by such and such an author, by such and such a cultural group, etc.) by sentences with high syntactic complexity could have been expressed with sentences of lower syntactic complexity, without any compensation. In this interesting interpretation, Wittgenstein's Thesis seems to me wrong, almost demonstrably so. I would, on the contrary, want to express and justify, if not really demonstrate, the following "Anti-Wittgensteinian Thesis": For most languages, and for all interesting (sufficiently rich) ones, there are things worth saying which cannot be expressed in sentences with a low degree of syntactic complexity, without a loss being incurred in other communicationally important respects.

Though a fuller justification will have to be postponed for another occasion, let me make here the following remarks. Consider one of the simplest calculi even invented by logicians, the so-called implicational propositional calculus [39, p. 140]. We are here interested only in its rules of formation but not in its axioms or theorems.

The rules of formation of one of the many formulations of this calculus are as follows: Its primitive symbols are the three improper symbols

$$],\qquad \supset,\qquad [$$

and the infinitely many proper symbols

$$p_1,\ p_2,\ p_3,\ \cdots\ .$$

Its rules of formation are just the following two:

F1. Each proper symbol is well-formed (wf).

F2. Whenever α and α are wf, so is $[\alpha \supset \beta]$

(with the understanding that nothing is wf unless it is so by virtue of F1 and F2). There exists no bound to the degree of nesting of the wf formulas of this calculus, as is obvious from the series of wf formulas

$$p_1,\ [p_1 \supset p_2],\ [p_1 \supset [p_2 \supset p_3]],\ [p_1 \supset [p_2 \supset [p_3 \supset p_4]]],\ \cdots\ .$$

It is less obvious, but can at any rate be rigorously proved, that for none of these formulas does there exist in the calculus another formula which is logically equivalent to it but has a lesser degree of nesting. (The term "logically equivalent" needs explanation in our context, but I shall never-

theless not provide it. For logicians the required explanation would be rather obvious, for non-logicians it would take too much time.) Wittgenstein's Thesis does not hold in this calculus.

Consider now the (logically utterly uninteresting) <u>conjunctional propositional calculus</u>, whose rules of formation are analogous to those of the implicational calculus, except that '$\supset$' is to be replaced by '$\wedge$' in both the list of improper symbols and F2. Here, too, it can be shown, by a somewhat more complicated argument, that for each n there exist wf formulas whose degree of nesting is higher than n such that they are not logically equivalent to any wf formula with a lesser degree of nesting.

But there exists the following interesting difference between the two calculi: The conjunctional calculus, as presented here, looks unduly complex. Since conjunction is "associative," i.e., since $[p_1 \wedge [p_2 \wedge p_3]]$ and $[[p_1 \wedge p_2] \wedge p_3]$ are equivalent, the brackets fulfill no semantically important function within the calculus and could as well have been omitted from the list of improper symbols, with a corresponding simplification in rule F2. In this version, all wf formulas would have had a degree of nesting 0, as can easily be verified! True enough, all formulas with at least two conjunction signs would have become syntactically ambiguous, but, in this particular calculus, syntactic ambiguity would not have entailed semantic ambiguity. Syntactic simplification could have been achieved, and in the most extreme fashion, without any semantic loss whatsoever!

This is by no means the case for the implicational calculus. Implication is not associative, so that the syntactic ambiguity introduced by omission of brackets would have entailed semantic ambiguity, a price no logician could possibly be ready to pay in this connection, though again all resulting formulas would have got a degree of nestedness 0.

(As for conjunctional calculus, as soon as it is combined with some other calculus, say the disjunctional calculus, omission of brackets would again entail semantic ambiguity, since, say, $[p_1 \wedge [p_2 \vee p_3]]$ and $[[p_1 \wedge p_2] \vee p_3]$ are not equivalent.)

For those of you who have heard of the so-called Polish bracket-free notation, let me add the following remark. One might have thought that the nesting (which in this particular case is also self-embedding) is due to the use of brackets for scoping purposes, in accordance with standard mathematical usage, since it seems that the brackets "cause" the branchings to be "inner" ones, and might therefore have cherished the hope that a bracket-free notation would eliminate, or at least reduce, nesting. But this hope is illusory. Inner branching, thrown out through the front door, would re-enter through the back door. With 'C' as the only improper symbol and F2 changed to: Whenever α and β are wf, so is C$\alpha\beta$, expansion of α (though not of β) causes inner branching. Notice further that in Polish notation calculi you cannot introduce syntactic ambiguity, harmless or harmful, even if you want to, by omitting symbols, since there are no special scoping symbols to omit.

As far as natural languages are concerned, the situation is much more confused. In speech, it seems that we can express distinctions of scope up to a degree of nesting of 3, anything beyond that becoming blurred, whereas in writing things are still worse, punctuation marks not being consistently used for scoping purposes and anyhow not being adequate for this task, with the result that syntactic ambiguities abound, which may or may not be reduced through context or background knowledge. Sometimes, when the resulting semantic ambiguity becomes intolerable, extraordinary measures are taken, such as using scoping symbols like parentheses in ways ordinarily reserved for mathematical formulas only, indentation at various depths, ad hoc abbrevaitions, etc.

Natural languages have many so-to-speak built-in devices for syntactic simplification. These devices, and their effectiveness, are badly in need of further study, after the extremely interesting beginnins by Yngve [30].

Certain "simplifications," beloved by editors who are out to split up involved sentences, may well turn out to be spurious and perhaps even down-

right harmful, in spite of appearances. An editor who rewrites an author's "Since p and q and r, therefore s." (where you have to imagine the letters p, q, r, and s replaced by sentences which on occasion will themselves have considerable syntactic complexity) by "p. q. r. Therefore s." is probably under the illusion that he has simplified something and therefore improved something. Now, he has doubtless replaced one long sentence with a degree of syntactic complexity of, say, n, by four shorter sentences, each with a degree of syntactic complexity of at most n-1, and has even used three words less for this purpose. But there is a price connected with this procedure, even a twofold one. First, the word "therefore" has become semantically much more indefinite. What for? "s, for r.", or " s, for q and r.", or "s, for p and q and r."? (And this might not be all. p will be preceded by other sentences, so that, at least from a purely syntactic point of view, it is totally indefinite how far back one has to go in the list of possible antecedents to s.) Secondly, even if the exact antecedent is settled, in order to understand the full content of the argument and to judge its validity, the reader (or listener) will have to recall, or re-read, the antecedent (which, so let us speculate, might have been removed into some larger, more permanent and less easily accessible storage than the immediate memory it was occupying during the syntactic processing), with the result that the overall economy of the "improvement" is, to say the least, very doubtful. There is at least a good chance that the total effort required of the receiver of the message will be higher in the case of the "split-up" sentence than with regard to the original sentence, though it might well be easier on the sender, had he wanted to express himself originally in this less definite way. (I used to teach geometry in high school and still remember the type of student who, when required to demonstrate a certain theorem, would start rattling off a list of congruences or inequalities, as the case might be, and finish with a triumphant "Therefore (or, "From this it follows that)...". And he was not even wrong. Because from his list, and in accordance with certain theorems already proved, his conclusion did indeed follow. Except that he left the task of finding out how, in detail, the conclusion followed from the premises, to the listeners,

including myself in that case, and provided no indication of the fact that he himself knew the details.)

An investigation, recently begun in Jerusalem, seems to lead to interesting results as to the mutual relationships between (semantic) equivalence among the sentences of a given formal system, the (syntactic) simplicity of these sentences and the existence of a recursive simplification function for this system. The results will be published in a forthcoming Technical Report. Let me only mention here one of the more significant results. (I hope to nobody's particular surprise.) The existence of a syntactic simplification algorithm is rather the exception, and the proof of such existence, if at all, will in general require that the system fulfill fairly tough conditions. The details, unfortunately, require a good knowledge of recursive function theory and shall therefore not be given here.

THIRD LECTURE: LANGUAGE AND SPEECH; THEORY VS. OBSERVATION IN LINGUISTICS

As already mentioned in the opening sentence of my first lecture, many of us believe that during the last few years we have gained valuable insights into the relationship between theory and observation in science. I myself have already tried on a few occasions to apply these insights to certain controversial issues of modern linguistics [40, 41]. I would now like to do the same with regard to the central term of linguistics, namely 'language' itself. As you will soon realize, this methodological point is of vital importance for the so-called "research methodology" in MT, and insufficient understanding of it has already caused superfluous controversies.

The term 'language' has, of course, been "defined" innumerably many times, but the fact that these definitions are usually mutually inconsistent, at least at first sight, has equally often been forgotten and neglected, so that seemingly contradictory statements about 'language' were usually interpreted as inconsistent statements about the same explicatum (in Carnap's terminology) rather than consistent statements about different explicata.

You will, for instance, find in the literature that language has often been treated as a <u>set of sentences</u> (or <u>utterances</u>, which two terms will not be distinguished for the moment). This, of course, is an abstraction from ordinary usage, and has been recognised as such. Leaving aside for our present purposes the discussion of how good and useful this abstraction is, let me point out that the characterisation can be understood (and has been understood) in at least the following five senses:

(1) A given set of utterances, such as recorded on a certain tape by so-and-so on such-and-such an occasion, or of inscriptions, found on such-and-such a tablet. Such sets are, of course, finite and most of them contain relatively few members. They can be, and sometimes are, represented as lists, under certain transcriptions. As a matter of fact, such sets are only exceptionally called 'languages', the more usual term being 'corpus'.

(2) The set of all utterances (spoken and/or written) made until July 1962, say, by the members of such and such a community during their lifetime until then. This set is certainly finite, too, but cannot, in general

be presented in list form and is rather indefinite, due to the indefiniteness of the term "community" and for dozens of other obvious reasons, such as those centering around idiolects, dialects, bilingualness, not to forget the vagueness of 'utterance' itself.

(3) The set of all utterances, past, present, and future, made by members of such a community. This set differs from that treated under (2) only in having a still greater degree of indeterminacy.

(4) The set of all "possible" utterances of a certain kind. The notion "possible" occurring in this characterization is notorious for its complexities and philosophical perplexities, and I trust I shall be forgiven if I don't go any deeper into this hornet's nest here. Under most conceptions, this set will turn out to be infinite.

(5) The set of all "sentences" (well-formed expressions, grammatical expressions, etc.). (For recent discussions of this and related hierarchies see, e.g., Quine [42] and Zipf [43].)

It is true, of course, that (1) is a subset of (2), which again is a subset of (3), but this is not the crucial point. Much more important is that the term 'utterance' occurring in their characterization changes its meaning in the transition from (3) to (4), becomes less observational and more theoretical. At the same time, there is a change from a concrete, physical, three- or four-dimensional entity, a "token," in Peirce's terminology, to an abstract entity, a "type." (When Paul and John say "I am hungry.", we have two members of the set (1), since they uttered two different "utterance-tokens," but only one member of the set (4), since these tokens are replicas of the same utterance-type.) The elements of set (5), finally, are so overtly theoretical that the term 'utterance' seemed definitely inappropriate for them, and I had to shift to the term 'sentence'. Though these two terms in ordinary usage, as well as in the usage of most linguists, are almost synonymous, I have already suggested once before [41] to distinguish artificially between them qua technical terms and use 'utterance' for observational entities and 'sentence' for theoretical ones (with the adjective 'possible' performing as a category-shifting modifier, an extremely important

and not fully analyzed semantical fact). That 'sentence' is ordinarily used in both these senses, as is 'word' and many other terms of this area, is, of course, one of the major sources of confusion and futile controversies.

Sets (2) and (3) have little linguistic importance. Because of their indefiniteness it is difficult to make interesting statements about them. Sets (4) and (5) — in all rigor I should have spoken about the classes of sets (4) and (5) — are by and large identical, at least under certain plausible interpretations of 'possible', the characteristic of (4) being what Carnap [44] called "quasi-psychologistic," while (5) is presumably characterized in an overtly and purely syntactical fashion.

In many linguistic circles, it has been standard procedure to make believe that linguists, in their professional capacity, are dealing with sets of type (1) (or of types (2) or (3)). This fiction gave their endeavor, so they believed, a closeness-to-earth, an operational solidity which they were anxious not to lose. In fact, they all, with hardly an exception, dealt with sets of types (4) or (5). All the talk about "corpora" was only lip-service. Today we know that no science worth its salt could possibly stick to observation exclusively. Whoever is out to describe and nothing else will not describe well. Theories are necessary evil. Though I don't think that it is necessary, or even helpful, to say that every description already contains theoretical elements — as some recent methodologists are fond of stressing — it must be said that theorophobia is a disease, fashionable as it might be. All scientific statements must surely be connected with observations, but this connection can, and must, be much more oblique than many methodological simplicists believe.

Returning from these generalities to our present problem of the relation between language and speech — with MT hovering in the back as a kind of proving ground — it should be superfluous to insist that the proper business of the theoretical linguist is to describe not the actual linguistic performance of some individual (or of so many individuals) — this "natural history" stage being of limited interest only — but his linguistic competence (or that of a certain community of individuals), to use a dichotomy that has recently been

much stressed by Miller and Chomsky [35]. Now competence is a disposition, perhaps even a higher-order disposition. To be a competent native speaker of English means not just to have performed in the past in a certain way, not even that he will (in all likelihood) perform in a certain way when presented with certain stimuli, but rather that one would perform, or would have performed (in all likelihood), in a certain way, were he to be presented (or had he been presented) with certain stimuli -- in addition to many other things. I know perfectly well that no competent English speaker will ever in his life be presented with a certain utterance consisting of a few billion words, say of the form "Kennedy is hungry, and Kruschev is thirsty, and De Gaulle is tired, ... , and Adenauer is old.", going over the whole present population of the world, but I know, and everybody else knows perfectly well, that were such a speaker, contrary to fact, to be presented with such an utterance, he would understand it as a perfect specimen of an English sentence.

There is no mechanical procedure to move from someone's performance to his competence, just as there is no mechanical procedure to move from any number of physical observations to a physical theory. But just as this fact does not free the physicist from his professional obligation to develop theories, so there is nothing to absolve the linguists from presenting theories of linguistic competence. Testing the validity of these theories will, again as in the other theoretical sciences, in general proceed not in any straightforward way but by standard indirect methods. That John is competent to understand a certain ten-billion-word sentence will not be tested by presenting John with a token of this sentence, but, as we all know, by entirely different, oblique methods. For the above sentence, for instance, it would suffice to find out that John understands such sentences as "Paul is hungry." and "David is thirsty." as well as that he has mastered the rule that whenever α and β are sentences, α followed by 'and' followed by β is a sentence. This latter finding might not be a very simple one or a very secure one, but we do often claim to have found out just such things.

One often hears, in certain philosophical circles as well as among people interested in applied linguistics, statements to the effect that natural lan-

guages have no grammar. These people are aware of the paradoxical character of such statements, but nevertheless insist that they are true, and even trivially so. Every grammar, so they say, determines a certain fixed, "static," set of sentences. But a natural language is a living affair, "dynamic," constantly in change, and it is utterly impossible that the set of sentences should coincide with the set of utterances, as it should for an adequate grammar. It should now be obvious where the fallacy lies in this argument: in the unthinking identification of sentences and utterances, and in the complete misunderstanding of the relation between theory and observation. It is as if one wanted to argue that natural gases obey no physical laws, since these laws apply only to the fictitious "ideal gases." (Incidentally, such statements have indeed been made by obscurantists at all times.) To understand the exact relationship between the laws of gases of theoretical physics and the behavior of real gases requires a lot of methodological sophistication, and no less should be expected for the understanding of the exact relationship between the grammatical rules of an artificial language and the utterances made by the members of the community speaking this language. Any naive identification will quickly result in paradox, futile discussions, and irrational distrust of theory.

That the question of the adequacy of a given grammar is much more complex than ordinarily assumed does not mean that this question is a pointless one. On the contrary, since there exists no simple criterion for deciding which of two propsed grammars is "better," more adequate than the other, the problem of finding <u>any</u> criterion, however partial and indirect, becomes of overwhelming importance. The fact is, of course, that extremely little is known here beyond programmatic declarations. We know that "grammatical" should not be identified with "comprehensible," nor is one of these concepts subsumed under the other, but neither are these two concepts incommensurable. In that connection we have the large complex of questions arising around degrees of grammaticalness, deviancy, oddness, and anomaly; all of vital importance to linguists and philosophers alike. Some of you know the valiant beginnings made toward an investigation of this problem by Chomsky, Ziff [43] and others, but it will, I hope, not deter you from following in their footsteps, if I

state, rather dogmatically, that these attempts are woefully inadequate, while admitting that I have nothing better to offer, for the moment.

As soon as it is understood that competence and performance are to be kept clearly apart, one will no longer be tempted to feel oneself obliged to impose upon, say, the English language a grammar which will not allow the generation of sentences of a higher degree of syntactic complexity than some small number, say 4, according to one or the other measures discussed in the previous lecture. True enough, "corresponding" utterances are not normally found in speech or writing, and if artificially produced will not be grasped unless certain artificial auxiliary means are invoked. These limitations of human performance are doubtless of vital importance; have to be clearly stated and investigated; and should, sooner or later, be backed up by some neurophysiological theory. They are of equal importance for the programming of machines which are charged with determining the syntactic structure of all sentences of any given text of a given language. That sentences of a high degree of complexity can be disregarded for this purpose, because of their extreme rarity or just plain non-occurrence, may allow an organisation of the computer's working space that could make all the difference between the economically feasible and the economically utopian. But in order to do all this, it is by no means necessary to impose these restrictions on the grammar of English as such. Nothing is gained, and much is lost. Not only will certain arbitrary-looking restrictions on the recursive generation rules have to be imposed, thereby increasing the complexity of the grammar to a degree that can hardly be estimated at present, but this procedure is self-defeating. It is done in the name of "sticking to the brute facts," but doing so in such a crude way will force the adherents of this approach to disregard other brute facts, such as that with the aid of certain auxiliary means, the syntactic structure of English word sequences of a degree of syntactic complexity of 5, or of 100 for that matter, will be perfectly grasped. Since these word sequences are not English sentences, according to the grammarians of performance, how come they are understood and what is the language they belong to?

This does not mean, of course, that restrictions of performance will not

reflect themselves in the grammar. I am convinced, e.g., that Professor Yngve has made a remark full of insight when he noticed and stressed the fact that by changing its mood from the active to the passive, the syntactic complexity of a given sentence can be reduced. And I have no objection to formulating this insight in the form that there exists a passive in English (and the same or other devices in other languages) in order to allow, among other things, the formulation of certain thoughts in sentences of a lower degree of complexity than would otherwise have been possible. But trying to obliterate the distinction between competence and performance, to say it for the last time, is only a sign of confusion and will breed further confusion. The sooner we get rid of these last traces of extreme operationalism, the better for all of us, including MT research workers.

In order to describe and explain the facts of speech exhaustively and revealingly, a full-fledged, formal theory of language is needed, among many other things. Philosophical prejudice aside, there is no particular merit in keeping this theory "close to the facts," in assuming that the rules of correspondence which connect the theory (in the narrower sense of the word) with observation will have a particularly simple form. Experience from other sciences should have taught us that such an assumption is baseless. Physics, e.g., has reached its present heights only because the free flight of fancy, "the free play of ideas," has not been fettered by a narrow conception of scientific methodology. True enough, the particular logical status of these rules of correspondence has still not been deeply enough investigated, and I fully understand the attitude of those who, for this reason, regard this whole business with suspicion, and are afraid that the free flight of fancy will reintroduce uncontrollable metaphysics into science in general and linguistics in particular. But I hope that the necessary controls will be developed and better understood in the future and that in the meantime one will manage somehow. Occasional metaphysical aberrations are probably less damaging in the long run than the curtailment of creative scientific imagination.

Let me stress, in this connection, that the extensive use of symbolism in the formulation of generative grammars has induced many linguists to accuse

the authors of these formulations of having lost all connection with empirical science and indulged instead in some mathematical surrogate. I hope that it is now perfectly clear that this accusation is baseless. A formal grammar of English is an empirical theory of the English language, and its symbolic formulation, while it increases its precision and therefore its testability, by no means turns it into a mathematical theory. When according to a certain grammar "Sincerity admires John." turns out not to be a (formal) sentence whereas this very sequence is considered by someone to be an (intuitive) sentence, then this grammar is to that degree inadequate to his intuitions. It should only be kept in mind that the determination of the intuitive sentencehood of "Sincerity admires John." is by no means such a straightforward affair of observation, experimentation and statistics as some people believe. The notion of "intuitive sentence" is highly theoretical itself (though without the benefit of a complete theory being formulated to back it up, which fact is, of course, the whole crux of this peculiar modifier 'intuitive'), and observations on utterances of people or their reaction to utterances alone will never settle in any clearcut way the question of the sentencehood of a particular word sequence. This is as it should be, and only wishful thinking and naive methodology make people believe otherwise. Confirmation and refutation of linguistic theories, as of theories in any other science, is not such a simple operation as one is taught to believe in high school. But the complexity of refutation does not make a linguistic theory empirically irrefutable and therefore does not turn it into a mathematical theory.

FOURTH LECTURE: WHY MACHINES WON'T LEARN TO TRANSLATE WELL

My arguments against the feasibility of high-quality fully-automatic translation can be assumed to be well known in this audience. I have gone through them often enough in lectures and publications. I also have the impression that, after occasionally rather strong initial negative reactions, a good number of people who have been active in the field of MT for some years tend more and more to agree with these arguments, though they might prefer a more restrained formulation. On the other hand, the number of research groups which have taken up MT as their major field of activity is still on the increase, and by now there is hardly a country left in Europe and North America which does not feature at least one such group, with Japan, China, India and a couple of South American countries joining them, for good measure. Though a certain amount of involvement in MT, and in particular in its theoretical aspects, is certainly helpful and apt to yield fresh insights into the workings of language, most of the work that is at present going on under the auspices of MT seems to me to be a wanton expenditure of research money that could be put to better use in other fields and, still worse, a deplorable waste of research potential.

The combined interest in MT is sometimes defended on the grounds that though it is indeed extremely unlikely that computers working according to rigid algorithms will ever produce high-quality translations, there still exists a possibility that computers with considerable learning ("self-organizing") abilities will be able through training and experience to improve their initial algorithms and thereby constantly improve their output until adequate quality is achieved. I myself mentioned the possibility in some prior publications but refrained from evaluating it, at that time regarding such an evaluation as premature [15,45].

During the last two years, however, while going through the pertinent literature once more and pondering over the whole issue of artificial intelligence, I came to more radical conclusions which I would like to expose and defend here. Today, I am convinced that even machines with learning abilities, as we know them today or foresee them according to known principles, will not be able to improve by much the quality of the translation output.

For this purpose, let us notice once more the obvious prerequisites for high-quality human translation. There are at least the following five of them, though deeper analysis would doubtless reveal more:

(1) competent mastery of the source language,

(2) competent mastery of the target language,

(3) good general background knowledge,

(4) expertness in the field,

and (5) intelligence (know-how).

(I admit, of course, that the last of these prerequisites, intelligence, is not too well defined or understood, and shall therefore have to use it with a good amount of caution.)

All this was surely common knowledge at all times, and certainly known to all of us "machine translations pioneers" a dozen years ago. I knew then that nothing corresponding to items (3) and (4) could be expected of electronic computers, but thought that (1) and (2) should be within their reach, and entertained some hopes that by exploiting the redundancy of natural language texts better than human readers usually do, we should perhaps be in a position to enable the computers to overcome, at least partly, their lack of knowledge and understanding. True enough, scientists (and almost everyone else) write their articles with a reader in mind who, in addition to having a good command of the language, has a general background knowledge of, say, college level, has so many years of study behind him in the respective field, and is intelligent enough to know how to apply these three factors when called upon to do so. But it could have been, couldn't it, that, perhaps inadvertently, they do introduce sufficient formal clues in their publications to enable a very ingenious team of linguists and programmers to write a translation program whose output, though produced by the machine without understanding, would be indistinguishable from a translation done out of understanding? After all, cases are known of human translations that were done under similar conditions and were not always recognised as such.

Well, it could have been so, but it just didn't turn out this way. For any given source language, there are countless sentences to which a competent

and not fully analyzed semantical fact). That 'sentence' is ordinarily used in both these senses, as is 'word' and many other terms of this area, is, of course, one of the major sources of confusion and futile controversies.

Sets (2) and (3) have little linguistic importance. Because of their indefiniteness it is difficult to make interesting statements about them. Sets (4) and (5) — in all rigor I should have spoken about the classes of sets (4) and (5) — are by and large identical, at least under certain plausible interpretations of 'possible', the characteristic of (4) being what Carnap [44] called "quasi-psychologistic," while (5) is presumably characterized in an overtly and purely syntactical fashion.

In many linguistic circles, it has been standard procedure to make believe that linguists, in their professional capacity, are dealing with sets of type (1) (or of types (2) or (3)). This fiction gave their endeavor, so they believed, a closeness-to-earth, an operational solidity which they were anxious not to lose. In fact, they all, with hardly an exception, dealt with sets of types (4) or (5). All the talk about "corpora" was only lip-service. Today we know that no science worth its salt could possibly stick to observation exclusively. Whoever is out to describe and nothing else will not describe well. Theoria sine necessaria est. Though I don't think that it is necessary, or even helpful, to say that every description already contains theoretical elements — as some recent methodologists are fond of stressing — it must be said that theorophobia is a disease, fashionable as it might be. All scientific statements must surely be connected with observations, but this connection can, and must, be much more oblique than many methodological simplicists believe.

Returning from these generalities to our present problem of the relation between language and speech — with MT hovering in the back as a kind of proving ground — it should be superfluous to insist that the proper business of the theoretical linguist is to describe not the actual linguistic performance of some individual (or of so many individuals) — this "natural history" stage being of limited interest only — but his linguistic competence (or that of a certain community of individuals), to use a dichotomy that has recently been

much stressed by Miller and Chomsky [35]. Now competence is a disposition, perhaps even a higher-order disposition. To be a competent native speaker of English means not just to have performed in the past in a certain way, not even that he will (in all likelihood) perform in a certain way when presented with certain stimuli, but rather that one would perform, or would have performed (in all likelihood), in a certain way, were he to be presented (or had he been presented) with certain stimuli -- in addition to many other things. I know perfectly well that no competent English speaker will ever in his life be presented with a certain utterance consisting of a few billion words, say of the form "Kennedy is hungry, and Kruschev is thirsty, and De Gaulle is tired, ... , and Adenauer is old.", going over the whole present population of the world, but I know, and everybody else knows perfectly well, that were such a speaker, contrary to fact, to be presented with such an utterance, he would understand it as a perfect specimen of an English sentence.

There is no mechanical procedure to move from someone's performance to his competence, just as there is no mechanical procedure to move from any number of physical observations to a physical theory. But just as this fact does not free the physicist from his professional obligation to develop theories, so there is nothing to absolve the linguists from presenting theories of linguistic competence. Testing the validity of these theories will, again as in the other theoretical sciences, in general proceed not in any straightforward way but by standard indirect methods. That John is competent to understand a certain ten-billion-word sentence will not be tested by presenting John with a token of this sentence, but, as we all know, by entirely different, oblique methods. For the above sentence, for instance, it would suffice to find out that John understands such sentences as "Paul is hungry." and "David is thirsty." as well as that he has mastered the rule that whenever α and β are sentences, α followed by 'and' followed by β is a sentence. This latter finding might not be a very simple one or a very secure one, but we do often claim to have found out just such things.

One often hears, in certain philosophical circles as well as among people interested in applied linguistics, statements to the effect that natural lan-

guages have no grammar. These people are aware of the paradoxical character of such statements, but nevertheless insist that they are true, and even trivially so. Every grammar, so they say, determines a certain fixed, "static," set of sentences. But a natural language is a living affair, "dynamic," constantly in change, and it is utterly impossible that the set of sentences should coincide with the set of utterances, as it should for an adequate grammar. It should now be obvious where the fallacy lies in this argument: in the unthinking identification of sentences and utterances, and in the complete misunderstanding of the relation between theory and observation. It is as if one wanted to argue that natural gases obey no physical laws, since these laws apply only to the fictitious "ideal gases." (Incidentally, such statements have indeed been made by obscurantists at all times.) To understand the exact relationship between the laws of gases of theoretical physics and the behavior of real gases requires a lot of methodological sophistication, and no less should be expected for the understanding of the exact relationship between the grammatical rules of an artificial language and the utterances made by the members of the community speaking this language. Any naive identification will quickly result in paradox, futile discussions, and irrational distrust of theory.

That the question of the adequacy of a given grammar is much more complex than ordinarily assumed does not mean that this question is a pointless one. On the contrary, since there exists no simple criterion for deciding which of two propsed grammars is "better," more adequate than the other, the problem of finding <u>any</u> criterion, however partial and indirect, becomes of overwhelming importance. The fact is, of course, that extremely little is known here beyond programmatic declarations. We know that "grammatical" should not be identified with "comprehensible," nor is one of these concepts subsumed under the other, but neither are these two concepts incommensurable. In that connection we have the large complex of questions arising around degrees of grammaticalness, deviancy, oddness, and anomaly; all of vital importance to linguists and philosophers alike. Some of you know the valiant beginnings made toward an investigation of this problem by Chomsky, Ziff [43] and others, but it will, I hope, not deter you from following in their footsteps, if I

state, rather dogmatically, that these attempts are woefully inadequate, while admitting that I have nothing better to offer, for the moment.

As soon as it is understood that competence and performance are to be kept clearly apart, one will no longer be tempted to feel oneself obliged to impose upon, say, the English language a grammar which will not allow the generation of sentences of a higher degree of syntactic complexity than some small number, say 4, according to one or the other measures discussed in the previous lecture. True enough, "corresponding" utterances are not normally found in speech or writing, and if artificially produced will not be grasped unless certain artificial auxiliary means are invoked. These limitations of human performance are doubtless of vital importance; have to be clearly stated and investigated; and should, sooner or later, be backed up by some neurophysiological theory. They are of equal importance for the programming of machines which are charged with determining the syntactic structure of all sentences of any given text of a given language. That sentences of a high degree of complexity can be disregarded for this purpose, because of their extreme rarity or just plain non-occurrence, may allow an organization of the computer's working space that could make all the difference between the economically feasible and the economically utopian. But in order to do all this, it is by no means necessary to impose these restrictions on the grammar of English as such. Nothing is gained, and much is lost. Not only will certain arbitrary-looking restrictions on the recursive generation rules have to be imposed, thereby increasing the complexity of the grammar to a degree that can hardly be estimated at present, but this procedure is self-defeating. It is done in the name of "sticking to the brute facts," but doing so in such a crude way will force the adherents of this approach to disregard other brute facts, such as that with the aid of certain auxiliary means, the syntactic structure of English word sequences of a degree of syntactic complexity of 5, or of 100 for that matter, will be perfectly grasped. Since these word sequences are not English sentences, according to the grammarians of performance, how come they are understood and what is the language they belong to?

This does not mean, of course, that restrictions of performance will not

reflect themselves in the grammar. I am convinced, e.g., that Professor Yngve has made a remark full of insight when he noticed and stressed the fact that by changing its mood from the active to the passive, the syntactic complexity of a given sentence can be reduced. And I have no objection to formulating this insight in the form that there exists a passive in English (and the same or other devices in other languages) in order to allow, among other things, the formulation of certain thoughts in sentences of a lower degree of complexity than would otherwise have been possible. But trying to obliterate the distinction between competence and performance, to say it for the last time, is only a sign of confusion and will breed further confusion. The sooner we get rid of these last traces of extreme operationalism, the better for all of us, including MT research workers.

In order to describe and explain the facts of speech exhaustively and revealingly, a full-fledged, formal theory of language is needed, among many other things. Philosophical prejudice aside, there is no particular merit in keeping this theory "close to the facts," in assuming that the rules of correspondence which connect the theory (in the narrower sense of the word) with observation will have a particularly simple form. Experience from other sciences should have taught us that such an assumption is baseless. Physics, e.g., has reached its present heights only because the free flight of fancy, "the free play of ideas," has not been fettered by a narrow conception of scientific methodology. True enough, the particular logical status of these rules of correspondence has still not been deeply enough investigated, and I fully understand the attitude of those who, for this reason, regard this whole business with suspicion, and are afraid that the free flight of fancy will reintroduce uncontrollable metaphysics into science in general and linguistics in particular. But I hope that the necessary controls will be developed and better understood in the future and that in the meantime one will manage somehow. Occasional metaphysical aberrations are probably less damaging in the long run than the curtailment of creative scientific imagination.

Let me stress, in this connection, that the extensive use of symbolism in the formulation of generative grammars has induced many linguists to accuse

the authors of these formulations of having lost all connection with empirical science and indulged instead in some mathematical surrogate. I hope that it is now perfectly clear that this accusation is baseless. A formal grammar of English is an empirical theory of the English language, and its symbolic formulation, while it increases its precision and therefore its testability, by no means turns it into a mathematical theory. When according to a certain grammar "Sincerity admires John." turns out not to be a (formal) sentence whereas this very sequence is considered by someone to be an (intuitive) sentence, then this grammar is to that degree inadequate to his intuitions. It should only be kept in mind that the determination of the intuitive sentencehood of "Sincerity admires John." is by no means such a straightforward affair of observation, experimentation and statistics as some people believe. The notion of "intuitive sentence" is highly theoretical itself (though without the benefit of a complete theory being formulated to back it up, which fact is, of course, the whole crux of this peculiar modifier 'intuitive'), and observations on utterances of people or their reaction to utterances alone will never settle in any clearcut way the question of the sentencehood of a particular word sequence. This is as it should be, and only wishful thinking and naive methodology make people believe otherwise. Confirmation and refutation of linguistic theories, as of theories in any other science, is not such a simple operation as one is taught to believe in high school. But the complexity of refutation does not make a linguistic theory empirically irrefutable and therefore does not turn it into a mathematical theory.

FOURTH LECTURE: WHY MACHINES WON'T LEARN TO TRANSLATE WELL

My arguments against the feasibility of high-quality fully-automatic translation can be assumed to be well known in this audience. I have gone through them often enough in lectures and publications. I also have the impression that, after occasionally rather strong initial negative reactions, a good number of people who have been active in the field of MT for some years tend more and more to agree with these arguments, though they might prefer a more restrained formulation. On the other hand, the number of research groups which have taken up MT as their major field of activity is still on the increase, and by now there is hardly a country left in Europe and North America which does not feature at least one such group, with Japan, China, India and a couple of South American countries joining them, for good measure. Though a certain amount of involvement in MT, and in particular in its theoretical aspects, is certainly helpful and apt to yield fresh insights into the workings of language, most of the work that is at present going on under the auspices of MT seems to me to be a wanton expenditure of research money that could be put to better use in other fields and, still worse, a deplorable waste of research potential.

The combined interest in MT is sometimes defended on the grounds that though it is indeed extremely unlikely that computers working according to rigid algorithms will ever produce high-quality translations, there still exists a possibility that computers with considerable learning ("self-organizing") abilities will be able through training and experience to improve their initial algorithms and thereby constantly improve their output until adequate quality is achieved. I myself mentioned the possibility in some prior publications but refrained from evaluating it, at that time regarding such an evaluation as premature [15,45].

During the last two years, however, while going through the pertinent literature once more and pondering over the whole issue of artificial intelligence, I came to more radical conclusions which I would like to expose and defend here. Today, I am convinced that even machines with learning abilities, as we know them today or foresee them according to known principles, will not be able to improve by much the quality of the translation output.

For this purpose, let us notice once more the obvious prerequisites for high-quality human translation. There are at least the following five of them, though deeper analysis would doubtless reveal more:

(1) competent mastery of the source language,

(2) competent mastery of the target language,

(3) good general background knowledge,

(4) expertness in the field,

and (5) intelligence (know-how).

(I admit, of course, that the last of these prerequisites, intelligence, is not too well defined or understood, and shall therefore have to use it with a good amount of caution.)

All this was surely common knowledge at all times, and certainly known to all of us "machine translations pioneers" a dozen years ago. I knew then that nothing corresponding to items (3) and (4) could be expected of electronic computers, but thought that (1) and (2) should be within their reach, and entertained some hopes that by exploiting the redundancy of natural language texts better than human readers usually do, we should perhaps be in a position to enable the computers to overcome, at least partly, their lack of knowledge and understanding. True enough, scientists (and almost everyone else) write their articles with a reader in mind who, in addition to having a good command of the language, has a general background knowledge of, say, college level, has so many years of study behind him in the respective field, and is intelligent enough to know how to apply these three factors when called upon to do so. But it could have been, couldn't it, that, perhaps inadvertently, they do introduce sufficient formal clues in their publications to enable a very ingenious team of linguists and programmers to write a translation program whose output, though produced by the machine without understanding, would be indistinguishable from a translation done out of understanding? After all, cases are known of human translations that were done under similar conditions and were not always recognized as such.

Well, it could have been so, but it just didn't turn out this way. For any given source language, there are countless sentences to which a competent

human translator will provide in a given target language many, sometimes very many, distinct renderings which will sometimes differ from each other only by minor idiosyncrasies, but will at other times be toto coelo different. The original sentence will very often be, as the standard expression goes, multiply ambiguous by itself, morphologically, syntactically, and semantically, but the competent human translator will render it, in its particular context, uniquely to the general satisfaction of the human reader. The translator will resolve these ambiguities out of the last three factors mentioned. Though it is undoubtedly the case that some reduction of ambiguity can be obtained through better attention to certain formal clues, and though it has turned out many times that what superficial thinking regarded as definitely requiring understanding could be handled through certain refinements of purely formal methods, it should by now be perfectly clear that there are limits to what these refinements can achieve, limits that definitely block the way to autonomous, high-quality, machine translation.

Could not perhaps computers with learning capacity do the job? Let me say rather dogmatically that a close study of one of the most publicized schemes for the mechanization of problem solving and a somewhat less detailed study of the whole field of Artificial Intelligence, has shown an amount of careless and irresponsible talk which is nothing short of appalling and sometimes close to lunatic. There is absolutly nothing in all this talk which shows any promise to be of real help in mechanizing translation. There is nothing to indicate how computers could acquire what the famous Swiss linguist de Saussure called, at the beginning of this century, the faculte de langage, an ability which is today innate in every human being, but which took evolution hundreds of millions of years to develop. Let nobody be deceived by the term "machine language" which may be suggestive for other purposes but which has turned out to be detrimental in the present context. Surely computers can manipulate symbols if given the proper instructions and they do it splendidly, many times quicker and safer than humans, but the distance from symbol manipulation to linguistic understanding is enormous, and loose talk will not diminish it.

Though certain electronic devices (such as perceptrons) have been built

which can be "trained" to perform certain tasks (such as pattern recognition) and indeed perform better after training than before, and though computers have been programmed to do certain things (such as playing checkers) and do these things better after a period of learning than before, it would be disastrous to extrapolate from these primitive exhibitions of artificial intelligence to something like translation. There just is no serious basis for such extrapolation. As to checkers, the definition of "legal move" is extremely simple and is, of course, given the computer in full. After a few years of work the inventor of the checker playing program [46] succeeded in formalizing a good set of strategies so that the training had nothing more to achieve than to introduce certain changes in the rank-ordering of these strategies. There never was any question of training the computer to discover the rules of checkers, or to expand an incomplete set of rules into a complete one, or to add new strategies to those given it beforehand. But some people do talk about letting computers discover rules of grammar or expand an incomplete set of such rules fed into it, by going over large texts and using "induction." But let me repeat, this talk is quite irresponsible and "induction " is nothing but a magic word in this connection. All attempts at formalizing what they believe to be inductive inference have completely failed, and inductive inference machines are pipe dreams even more than autonomous translation machines.

Now children do learn, as we all know, their native language up to an almost complete mastery of its grammar by the time they are four or five years old. But by the time they reach this age, they have heard (and spoken) surely no more than a few hundred thousand utterances in their native language (only a part of which are good textbook specimens of grammatical sentences). If they succeeded in mastering the grammar, apparently "by induction" from these utterances, why shouldn't a computer be able to do so? Even if we add the fact that these children were also told that so many word sequences were not grammatical sentences — whatever the form was by which they were given these pieces of instruction —,could not the same procedure be mirrored for computers? Well, the answer to these two questions can be nothing but an uncompromising No. The children are able to perform as splendidly as they do because, in

addition to the training and learning, their brain is not a tabula rasa general purpose computer but a computer which, after all those hundreds of thousands of years of evolution mentioned before, is also special purpose structured in such a way that it possesses the unique faculté de langage which makes it so different from the brain of mice, monkeys, and machines. The fact that we know close to nothing about this structure does not turn the previous statement into a scholastic truism.

Years of most patient and skillful attempts at teaching monkeys to use language intelligently succeeded in nothing better than making them use four single words with understanding, and monkeys' brains are in many respects vastly superior to those of computers. True enough, computers can do many things better than monkeys or humans, computing for instance, but then we know the corresponding algorithms, and know how to feed them into the computer. In some cases we know algorithms which, when fed into the computer, will enable it to construct for itself computing algorithms out of other data and instructions that can be fed into it. But nothing of the kind is known with respect to linguistic abilities. So long as we are unable to wire or program computers so that their initial state will be similar to that of a newborn human infant, physically or at least functionally, let's forget about teaching computers to construct grammars.

Let me now turn to the first two items. What is the outlook for computers to master a natural language to approximately the same degree as does a native speaker of such a language? And by "mastering a language" I now mean, of course, only a mastery of its grammar, i.e. vocabulary, morphology, and syntax, to the exclusion of its semantics and pragmatics. Until recently, I think that most of us who dealt with MT at one time or another believed that not only was this aim attainable, but that it would not be so very difficult to attain it, for the practical purpose at hand. One realized that the mechanization of syntactic analysis, based on this mastery, would lead on occasion to multiple analyses whose final reduction to a unique analysis would then be relegated to the limbo of semantics, but did not tend to take this drawback very seriously. It seems that here, too, a more sober appraisal

of the situation is indicated and already is gaining ground, if I am not mistaken. More and more people have become convinced that the inadequacies of present methods of mechanical determination of syntactic structure, in comparison with what competent and linguistically trained native speakers are able to do, are not only due to the fact that we don't know as yet enough about the semantics of our language — though this is surely true enough — but also to the perhaps not too surprising fact that the grammars which were in the back of the minds of almost all MT people were of too simple a type, namely of the so-called immediate constituent type, though it is quite amazing to see how many variants of this type came up in this connection.

Leaving aside the question of the theoretical inadequacy of immediate constituent grammars for natural languages, the following fact has come to the fore during the last few years: If one wants to increase the degree of approximate practical adequacy of such grammars, one has to pay an enormous price for this, namely a proliferation of rules (partly, but not wholly, caused by a proliferation of syntactic catagories) of truly astronomic nature. The dialectics of the situation is distressing: the better the understanding of linguistic structure, and greater our mastery of the language — the larger the set of grammatical rules we need to describe the language, the heavier the preparatory work of writing the grammar, and the costlier the machine operations of storing and working with such a grammar.

It is very often said that our present computers are already good enough for the task of MT and will be more than sufficient in their next generation, but that the bottleneck lies mostly in our insufficient understanding of the workings of language. As soon as we know all of it, the problem will be licked. I shall not discuss here the extremely dubious character of this "knowing all of it," but only point out that the more we shall know about linguistic structure, the more complex the description of this structure will become, so long as we stick to immediate constituent grammars. It is known that in some cases transformational grammars are able to reduce the complexity of the description by orders of magnitude. Whether this holds in

general remains to be seen, but the time has come for those interested in the mechanical determination of syntactic structure, whether for its own sake, for MT or for other applications, to get out of the self-imposed straitjacket of immediate constituent grammars and start working with more powerful models, such as transformational grammars.

Let me illustrate by just one example: one of the best programs in existence, on one of the best computers in existence, recently needed 12 minutes (and something like $100 on a commercial basis) to provide an exhaustive syntactic analysis of a 35-word sentence [47]. I understand that the program has been improved in the meantime and that the time required for such an analysis is now closer to one minute. However, the output of this analysis is multiple, leaving the selection of the single analysis, which is correct in accordance with context and background, to other parts of the program or to the human posteditor. But there are other troubles with using immediate constituent grammars only for MT purposes. In his lecture to this Institute, Mr. Gross gave an example of a French sentence in the passive mood which could be translated into English only by ad hoc procedures so long as its syntactic analysis is made on an immediate constituent basis only. The translation into English is straightforward as soon as the French sentence is first detransformed into the active mood. A grammar which is unable to provide this conversion, besides being scientifically unsatisfactory, will increase the difficulties of MT.

In the time left to me I would like to return to what is perhaps the most widespread fallacy connected with MT, the fallacy I call, in variation of a well known term of Whitehead, The Fallacy of Misplaced Economy. I refer to the idea that indirect machine translation through an <u>intermediate language</u> will result in considerable to vast economies over direct translation from source to target language, on the obvious condition that should MT turn out to be feasible at all, in some sense or other, many opportunities for simultaneous translation from one source language into many target languages (and vice versa) will arise. I already once before discussed both the attractiveness of this idea and the fallaciousness of the reasoning behind it. Let

me therefore discuss here at some length only what I regard to be the kernel of the fallacy.

The following argument has great prima facie appeal: Assume that we deal with 10 languages, and that we are interested in translating from each language into every other, i.e., altogether 90 translation pairs. Assume, for simplicity's sake, that each translation algorithm -- never mind the quality of the output -- requires 100 man-years. Then the preparation of all the algorithms will require 9000 man-years. If one now designates one of these languages as the pivot-language, then only 18 translation pairs will be needed, requiring 1800 man-years of preparation, an enormous saving. True enough, translation time for any of the remaining 72 language pairs will be approximately doubled, and the quality of the output will be somewhat reduced, but this would be a price worth paying. (In general, the argument is presented with some artificial language serving as the pivot. Though this move changes the appeal of the argument for the better -- since this artificial pivot language is supposed to be equipped with certain magical qualities -- as well as for the worse -- since the number of translation algorithms now increases to 20--I don't think that thereby the substance of the following counterargument is weakened.) However, in order to counteract even this deterioration, let us double our effort and spend, say, 200 man-years on the preparation of the algorithms for translating to and from the pivot language. We would still wind up with no more than 3600 man-years of work, vs. the 9000 originally needed. Well?

The fallacy, so it seems to me, lies in the following: the argument would hold if the preparation of the 90 algorithms were to be done independently and simultaneously by different people, with nobody learning from the experience of his co-workers. This is surely a highly unrealistic assumption. If preparing the Russian-to-English and German-to-English algorithms were to take 100 man-years each, when done this way, there can be no doubt that preparing the German-to-English algorithm after completion (or even partial completion) of a successful Russian-to-English algorithm will take much less time, perhaps half as much. The next pair, say Japanese-to-English, will take still less time, etc. All these figures being utterly arbitrary, I don't think we should

go on bothering about the convergence of this series. Though we might still wind up with a larger time needed for the preparation of the 90 than of the 18 "double precision" algorithms, it is doubtful, to say the least, whether the overall quality/preparation-time/translation-time balance would be in favor of the pivot language approach.

Add to this the fact that 100 man-years would be enough, by assumption, to start a working MT outfit along the direct approach, whereas 400 man-years will be needed even to start translating the first pair along the indirect approach, and the initial appeal of the intermediate language idea should completely vanish, when judged from a practical point of view. As to its speculative impact, enough has been said on other occasions.

I think it is my duty to state at the end of this lecture series where all this leaves us. Autonomous, high-quality machine translation between natural languages according to rigid algorithms may safely be considered as dead. Such translation on the basis of learning abilities is still-born. Though machines could doubtless provide a great variety of aids to human translation, so far in no case has economic feasibility of any such aid been proven, though the outlook for the future is not all dark. So much for the debit side. On of credit side of the past MT efforts stands the enormous increase of interest which has already begun to pay off not only in an increased understanding of language as such, but also in such applications as the mechanical translation between programming languages. But this could already be a topic for another Institute.

REFERENCES

[1] Harris, Z. S. Methods in structural linguistics. Chicago, Univ. of Chicago Press, 1951.

[2] Hockett, C. F. Two models of grammatical description. Word, vol. 10 (1954), 210-231 (reprinted as Ch. 39 in Readings in linguistics (M. Joos, ed.), Washington D. C., American Council of Learned Societies).

[3] Hjelmslev, L. Prolegomena to a theory of language (tr. by F. J. Whitfield), Baltimore, Waverly Press, 1953.

[4] Uldall, H. Outline of glossematics. Copenhagen, Nordisk Sprog- og Kulturforlag, 1957.

[5] Carnap, R. The logical syntax of language. New York, Harcourt, Brace & Co., 1937.

[6] Ajdukiewicz, K. Die syntaktische Konnexität. Studia Philosophica, vol. 1 (1935), 1-27.

[7] Post, E. L. Formal reductions of the general decision problem. American Journal of Mathematics, vol. 65 (1943), 197-215.

[8] Davis, M. Computability and unsolvability. New York, McGraw-Hill, 1958.

[9] Curry, H. B. and Feys, R. Combinatory logic. Amsterdam, North-Holland Publishing Co., 1958.

[10] Chomsky, N. Syntactic Structures. s'Gravenhage, Mouton & Co., 1957.

[11] Bar-Hillel, Y. A quasi-arithmetical notation for syntactic description. Language, vol. 29 (1953), 47-58.

[12] Lambeck, J. The mathematics of sentence structure. American Mathematical Monthly, vol. 65 (1958), 154-170.

[13] Lambeck, J. Contributions to a mathematical analysis of the English verb-phrase. Journal of the Canadian Linguistic Association, vol. 5 (1959), 83-89.

[14] Lambeck, J. On the calculus of syntactic types. Twelfth Symposium in Applied Mathematics (R. Jakobson, ed.), Providence, R. I., American Mathematical Society, 1961.

[15] Bar-Hillel Y. The present status of automatic translation of languages, Appendix II, in Advances in Computers, Vol. I (F. L. Alt, ed.), New York, Academic Press, 1960.

[16] Bar-Hillel, Y., Gaifman, C., and Shamir, E. On categorial and phrase-structure grammars. Bulletin of the Research Council of Israel, vol. 9F (1960), 1-16.

[17] Rabin, M. O., and Scott, D. Finite automata and their decision problems, IBM Journal of Research and Development, vol. 3 (1959), 115-125.

[18] Bar-Hillel, Y., and Shamir, E. Finite-state languages: Formal representations and adequacy problems. Bulletin of the Research Council of Israel, vol. 8F (1960), 155-166.

[19] Bar-Hillel, Y., Perles, M., and Shamir, E. On formal properties of simple phrase-structure languages. Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung, vol. 14 (1961), 143-172.

[20] Post, E. A variant of a recursively unsolvable problem. Bulletin of the American Mathematical Society, vol. 52 (1946), 264-268.

[21] Kleene, S. C. Representation of events in nerve nets and finite automata. Automata Studies (C. E. Shannon and J. McCarthy, eds.), Princeton Univ. Press, 1956.

[22] Ginsburg, S., and Rice, H. G. Two families of languages related to ALGOL. TM-578/000/01, SDC, Santa Monica, Cal., July 1961.

[23] Ginsburg, S., and Rose, G. F. Operations which preserve definability in languages. SP-511, SDC, Santa Monica, Cal., October 1961.

[24] Shamir, E. On sequential languages. Applied Logic Branch, Technical Report No. 7 (prepared for the Office of Naval Research, Information Systems Branch), Hebrew University, Jerusalem, Israel, November 1961.

[25] Hays, D. G. Grouping and dependency theories. P-1910, Rand Corporation, Santa Monica, Cal., 1960.

[26] Lecerf, Y. and Ihm, P. Eléments pour une grammaire générale des langues projectives. Rapport GRISA, No. 1, 11-29, 1960.

[27] Tesnière, L. Eléments de syntaxe structurale. Paris, Klincksieck, 1959.

[28] Gaifman, C. Dependency systems and phrase-structure systems. P-2315, Rand Corporation, Santa Monica, Cal., 1961.

[29] Chomsky, N. Formal properties of grammars, Handbook of Mathematical Psychology (in press).

[30] Yngve, V. H. A model and an hypothesis for language structure. Proceedings of the American Philosophical Society, vol. 104 (1960), 444-466.

[31] Yngve, V. H. The depth hypothesis. Twelfth Symposium in Applied Mathematics (R. Jakobson, ed.), Providence, R. I., American Mathematical Society, 1961.

[32] Myhill, J. Linear bounded automata. Wright Air Development Division, Technical Note 60-165, 1960.

[33] McNaughton R. The theory of automata, a survey. Advances in Computers, Vol. II (Franz L. Alt, ed.), New York, Academic Press, 1962.

[34] Flesch, R. A new readability yardstick. Journal of Applied Psychology, vol. 32 (1948), 221-233.

[35] Chomsky, N. and Miller, G. A. Finitary models of language users. Handbook of Mathematical Psychology (in press).

[36] Chomsky, N. On certain formal properties of grammars, Information and Control, vol. 2 (1953), 133-167.

[37] Chomsky, N. On the notion "rule of grammar", Twelfth Symposium in Applied Mathematics (R. Jakobson, ed.), Providence, R. I., American Mathematical Society, 1961.

[38] Chomsky, N. Explanatory models in linguistics. Logic, Methodology and Philosophy of Science: Proceedings of the 1960 International Congress (E. Nagel, P. Suppes, and A. Tarski, eds.), Stanford, Cal., Stanford University Press, 1962.

[39] Church, A. Introduction to mathematical logic, vol. I. Princeton University Press, 1956.

[40] Bar-Hillel, Y. Recursive definitions in empirical sciences. Proceedings of the Eleventh International Congress of Philosophy, vol. 5 (1953), Brussels, 160-165.

[41] Bar-Hillel, Y. Three remarks on linguistics fundamentals. Word, vol. 13 (1957), 323-335.

[42] Quine, W. V. From a logical point of view. Harvard University Press, 1953.

[43] Ziff, P. Semantic analysis. Ithaca, New York, Cornell Univ. Press, 1960.

[44] Carnap, R. Logical foundations of probability. University of Chicago Press, 1950.

[45] Bar-Hillel, Y. The future of machine translation. Freeing the Mind, Articles and Letters from Times Literary Supplement during March-June 1962, 32-37.

[46] Samuel, A. L. Programming computers to play games. Advances in Computers Vol. I (F. L. Alt, ed.), New York, Academic Press, 1962.

[47] Kuno, S. and Oettinger, A. G. Multiple-path syntactic analyzer. Proceedings of the IFIP Congress 1962 (in press).