[From: *Trends in European and American Linguistics 1930-1960*,
ed.Christine Mohrmann, Alf Sommerfelt and Joshua Whatmough
(Utrecht: Spectrum, 1963)]

MATHEMATICAL LINGUISTICS

WARREN PLATH

1. INTRODUCTION

Since the time of the Eighth International Congress of Linguists in 1957 there has been a considerable growth of interest and activity in the field of mathematical linguistics, both in America and in Europe. This activity has recently spread to the Far East as well, with the establishment in Japan of a journal[1] and a society[2] for mathematical linguistics. In his report to the 1957 Congress, Professor Joshua Whatmough mentioned the initiation of the Seminar in Mathematical Linguistics at Harvard University two years previously – the first appearance of the subject in an academic curriculum. The intervening years have brought a significant expansion in terms of individual research and formal courses offered in mathematical linguistics, not only at Harvard[3], but at a number of other universities in Europe and the United States. These other academic centers include the Massachusetts Institute of Technology, the University of Michigan and the University of Pennsylvania in the United States, the University of Bonn in Germany, and also Moscow and Leningrad Universities in the Soviet Union. Of late there have been indications that mathematical linguistics is gaining still wider acceptance as an academic discipline. Courses in quantitative linguistics have been introduced at Indiana University and at the University of California, and in 1958 the Soviet Minister of Education issued an order that the rectors of Moscow, Leningrad, Gorky, Saratov and Tomsk Universities introduce elective courses in machine translation and mathematical linguistics for students of mathematics and philology.

It is the purpose of this paper to review the recent developments in the field of mathematical linguistics, with emphasis on some of the major theoretical concepts and fields of application whose outlines

---

1) *Mathematical Linguistics* (1957-).

2) The Mathematical Linguistics Society of Japan.

3) There is an annual volume of seminar papers on file in the Harvard College Library.

are emerging from the large and varied body of literature which has appeared in this area. Due to the limited space available for this study, certain valuable individual contributions will inevitably have to be neglected, as will a number of important fields[1] closely related to mathematical linguistics. It is hoped, however, that in such cases the accompanying list of bibliographical references, while by no means exhaustive, will serve as a guide to those who wish to attain a more complete view of a given topic.

Much of the development of linguistics has been based on the success and power of discrete models[2] in the practical representation and theoretical description of the formal properties of natural languages. Although individual language models, ranging from the orthographical system of English to recently constructed descriptive grammars of exotic languages, have frequently been strongly criticized for various weaknesses, it has rarely been suggested that discrete models as a class are inadequate to the task. On the contrary, most criticisms of this type embody the suggestion that another discrete model with more desirable properties (from the critic's point of view) be substituted for the original one.

If language models are considered as abstract systems of discrete elements, a variety of mathematical concepts and techniques, ranging from elementary ideas of number to complex logical, statistical and set-theoretical operations, may be applied to them. However, the idea that the mere attachment of numbers or mathematical operations to the elements of a system makes statements expressed in terms of the system either more 'exact' or more 'scientific' is completely erroneous. What has to be demonstrated is that the new system so obtained is in some sense a more satisfactory model than the original system, either in that it makes possible the formulation of simpler or more general theoretical statements about certain aspects of the modelled domain[3] or in that operations on the model can give new insights into the results of corresponding operations in the modelled domain. One of the greatest dangers involved in the construction of mathematical models of language, particularly quantitative ones, is

---

1) Particularly the fields of speech analysis and synthesis. The reader is referred in this connection to the *Journal of the Acoustical Society of America,* in which numerous papers relating to these topics have appeared.

2) For a discussion of the concept of a model, cf. Oettinger (1957). Cf. also Joos (1950), who characterizes linguistics as a discrete mathematics.

3) Chomsky (1957) has emphasized the importance of criteria of this type in connection with the evaluation of the relative merits of various types of grammars.

that indiscriminate introduction of mathematical apparatus inevitably leads to the generation of meaningless and misleading results. It should therefore be clear that the prerequisites for significant contributions to linguistics from this quarter include not only knowledge of the pertinent branches of mathematics, but, in addition, a thorough comprehension of the nature of the linguistic problems to which the mathematical methods are to be applied.

Elementary mathematical procedures based on counting were applied to the study of language even in ancient times.[1] In recent years, simple enumerations of symbol occurrences have largely given way to studies in which techniques of mathematical statistics and information theory have been applied to the data obtained from counts.[2] These latter methods have been employed with greatest success in the study of large-scale language phenomena, the applications ranging from the investigation of certain over-all regularities in the vocabulary distribution of individual languages and authors to the development of measures of the degree of genetic relationship between pairs of natural languages. Some of the major developments in this field of 'quantitative' or 'statistical' linguistics will be discussed at greater length in Part 2.

Much of modern linguistics is concerned with the construction of models which represent languages in terms of the possible structural configurations and interrelations of elements defined on a number of different levels (phonology, morphology, syntax). The mathematics of modern algebra appears to be particularly well suited to the rigorous development of such descriptive systems, and a considerable amount of research is currently being conducted along these lines.[3] In Part 3, which will be devoted to structural models, it will be shown that quantitative methods, while not of primary importance in this latter area, can play a meaningful subsidiary role in some structural models of language.

Part 4, the concluding section of this paper, will include a brief consideration of some of the important practical applications of mathematical linguistics. The current intensive activity in such fields as machine translation, information analysis and retrieval, and the

---

1) Cf. Whatmough(1957).

2.) For an extensive critical bibliography of statistical linguistics, the reader is referred to Guiraud (1954).

3) The reader's attention is called to the Symposium on the Structure of Language and its Mathematical Aspects (1960), at which a number of papers relating to this and other areas of mathematical linguistics were presented.

design of automatic programming languages for computers has already proved to be a great stimulus to the development of mathematical linguistics and will in all probability exert a strong influence on the future of linguistics as a whole.

2. STATISTICAL MODELS OF LANGUAGE

‘MACROLINGUISTIC’ MODELS. Probably the most familiar and widely-discussed formula which has been proposed in the area of statistical linguistics is the one commonly known as Zipf’s Law (Zipf, 1949): $r \bullet f = C$.[1] The formula states essentially that if the words in a long text are ranked in order of decreasing frequency of occurrence in the text, so that the most frequent word has the rank $r = 1$, the next most frequent has rank $r = 2$, and so forth, then the product of the rank r and the frequency f for any word in the text will be approximately the same constant C, where C depends on the length of the text.

The regularity of this relationship, verified by Zipf for texts from a wide variety of languages, has attracted the attention of a number of scholars, since it has seemed to hold a potential key to some general principle of linguistic behavior. Zipf himself interpreted the data as evidence of the operation of a fundamental law of human behavior which he called the ‘Principle of Least Effort’, by analogy to the principle of least action in physics. However, this interpretation has gained little acceptance as an explanatory theory for rank-frequency distributions, since the vagueness of the underlying principle has made it an unsatisfactory basis for the construction of mathematical models of text generation which can be tested for consistency with the observed data.

During the past decade, a considerable amount of effort, centered about the work of Mandelbrot, has been devoted to attempts at ‘explaining’ rank-frequency distributions in terms of mathematical models based on hypotheses other than the one proposed by Zipf. Mandelbrot (1957) has presented the results of his investigations within the broad theoretical framework of ‘macrolinguistics’. He conceives of this proposed discipline as a new branch of linguistics which is to be concerned exclusively with the description and study, by statistical

---

[1] Actually 1929; see *Harvard Studies in Classical Philology* 40, 1929, p. 89 where Zipf wrote $XY = n$ (afterwards $r \bullet f = C$) and gave the same graph. This was in his Ph. D. thesis in Linguistics (Harvard 1929, cf. pp. 2-3, n. 1); see also his *Selected Studies* (H.U.P. 1932), p. 27.

means, of large-scale linguistic phenomena. The position of macrolinguistics in relation to microlinguistics (grammar) should, according to Mandelbrot, be closely analogous to that of thermodynamics in relation to the mechanics of individual molecules of a gas: the description on the macroscopic level, while consistent with the microscopic behavior described by a grammar or by the laws of mechanics, ignores the details of behavior on the lower level. Through such simplification, the macroscopic approach of thermodynamics, although necessarily only an incomplete description of the behavior of gases, has been extremely useful in that it has easily led to computational results which it would have been virtually impossible to obtain from consideration of the motion of individual molecules. Mandelbrot has suggested that macrolinguistics may provide a similarly valuable tool for describing the gross features of large masses of textual material for which a complete, detailed grammatical treatment might be unmanageably extensive and complex.

In his investigations of rank-frequency distributions from the ‘macrolinguistic’ point of view, Mandelbrot has proposed that Zipf’s original formula be modified to correspond more closely to the observed data by the introduction of two new parameters, ρ and B, yielding what he terms the Canonical Law:[1] $p_r = P(r + \rho)^{-B}$. Here r is the rank, as before, p_r is the relative frequency of the word of rank r and P, ρ, and B are constants for a given text; ρ provides a correction for the words of low rank, and $-B$ (which in the Zipf formula is equal to -1) corresponds to the slope of a plot of p_r versus r on bilogarithmic graph paper.

Mandelbrot has succeeded in deriving the Canonical Law mathematically from two different theoretical models of the generation of texts. According to the first and simpler model, the words of a text are assumed to be generated letter by letter by means of a finite-state Markov process[2] in which each symbol, including the space, has some fixed probability of occurrence. The production of a text by a probabilistic model of this type, with random placement of the space, results in a distribution of word frequencies which follows the Canonical Law with B greater than 1.[3] Mandelbrot’s second model was developed in terms of an analogy with thermodynamics: he determines mathematic-

---

1) The Canonical Law is stated here in the form in which it appears in some of Mandelbrot’s earlier articles, viz. Mandelbrot (1954, 1955).

2) Finite-state Markov processes are discussed more fully in Part 3 in the section *Models for Sentence Synthesis and Syntactic Description.*

3) Mandelbrot (1957), pp. 36-37.

ally the ‘most probable state’ of a text, subject to the restrictions that all words be separated by a space during the decoding process and that the cost of optimal decoding (that is, a decoding system in which the shortest series of operations is assigned to the most frequent word, and so forth) be fixed. This determination involves the maximization of the entropy (in the sense employed by Shannon)[1] associated with the distribution of word probabilities and once again leads to the result that the distribution of word frequencies follows the Canonical Law, this time with no restriction on the values of *B*.

On the basis of the second model, which he evidently prefers, Mandelbrot draws several conclusions, among them that words are the basic units of texts and that information theory has an intrinsic significance for linguistics. Other workers who have dealt with this problem have objected that there is no need to make the strong assumptions involved in the information-maximizing model, and that the conclusions based upon it are therefore not necessary consequences of the data. Miller and Newman (1958) have made a particularly good case for Mandelbrot’s first model by pointing out that long strings of letters uninterrupted by spaces are unlikely and that there is necessarily a greater variety of longer words, so that it would be very difficult to avoid an essentially random placement of the space over long texts.[2]

Among the other models which have been proposed is that of Simon (1955), who treats the formation of a text as a stochastic ‘birth process’ and from this assumption derives a distribution function relating the number of word types of a given frequency to their frequency of occurrence. Mandelbrot (1959) has stated that Simon’s derivation is essentially circular, and that consequently even though the function can be made to fit the data reasonably well, this gives no positive indication that anything like a ‘birth process’ is involved. Belevitch (1959) and Somers (1959) have suggested that the assumption that the logarithm of the relative frequencies of words is normally distributed is sufficient to lead to rank-frequency distributions of the Zipf type. By means of a first and then a second order Taylor series approximation to a truncated normal distribution, Belevitch derives Zipf’s Law and Mandelbrot’s Canonical Law, respectively. He maintains that no special assumptions are needed to explain a normal distribution, although he in fact has made a very strong assumption in stating that

---

1) Cf. the discussion in the section *Information-theoretic Models* further on in Part 2 of this paper.

2) Cf. also Miller (1957) for his original statement of this argument.

the logarithm of a probability is the natural variable to consider in statistical linguistics.

Since a variety of statistical models, including the very simple random space model, lead to rank-frequency distributions of the classical form described by Zipf and Mandelbrot, it should not be surprising that such distributions are found to hold for the great majority of long texts. It appears likely, therefore, that the regularity of rank-frequency distributions for texts will in itself give little insight into basic linguistic processes, and that the validity of any of the more complicated models proposed will have to be tested using data from other sources, such as psychological tests. This does not indicate that the effort expended on this problem has been wasted, however, for the establishment of the fact that the Canonical Law with $B = 1$ is so often satisfied suggests the use of such distributions as a yardstick for measuring significant deviations in linguistic behavior. Mandelbrot has already proposed that the parameter B (actually in the form $1/B$) may provide a useful measure of vocabulary efficiency with possible applicability to the measurement of intelligence and the detection of certain pathological mental conditions. Thus the ‘macrolinguistic’ approach appears to be making its chief contributions in a region common to psychology and linguistics, rather than in the more traditional area of grammar. This is a fairly natural consequence of the use of models which, by their very definition, do not include the details which form the essence of grammatical description.

STATISTICS OF STYLE AND AUTHORSHIP. In contrast to the emphasis, in the work just discussed, on the general similarity of rank-frequency distributions, the study of text statistics has been approached by certain other scholars with the radically different aim of finding statistical measures which may best serve to bring out the stylistic differences among the works of various authors.[1] Members of this latter group have attempted to bring quantitative criteria to bear on such problems as those of disputed authorship, the relative chronology of the works of a single author, and the description of literary style in general - ostensibly in the hope that such criteria may lead to more objective and reliable judgments than have been possible previously. Everyday usage of the term 'literary style' is very flexible, ranging from the characterization of a single work (or part of a work) by a

---

[1] Efforts which have been made by Luhn (1958) and others to apply similar techniques in systems for automatic literature abstracting will be mentioned in Part 4.

single author to that of all the works of a given author, group of authors or historical period. Descriptions of style often involve considerations of both form and content; they may range from a discussion of sound patterning in poetry to an examination of the arrangement of factual material in an entire prose work. A concept of such vagueness and generality clearly cannot be taken as a whole and 'reduced to mathematics' directly with any meaningful results. What has been done in a few cases, however, is to single out a particular stylistic attribute (almost invariably a formal one) to be treated quantitatively, usually on the basis of the relative frequency of occurrence of linguistic forms of a specified type.[1]

The classical example of the latter approach is Yule's work (Yule, 1944) on the statistics of literary vocabulary, which developed from his interest in the controversy over the authorship of the *De Imitatione Christi.* Starting with a study of the frequency distributions of nouns in the disputed work and in those of the two most likely candidates for its authorship, Thomas à Kempis and Jean Gerson, Yule encountered a series of basic methodological problems which led him to extend the scope of his investigations considerably. His work is particularly illuminating in its emphasis on two important points: first, the great difficulties involved in sampling data of this type, and the precautions necessary to insure that the sample be unbiased; second, the importance of finding statistical measures which are independent of sample size, so that results obtained from texts of different lengths may be compared directly. Yule's chief contribution in the latter connection is his introduction of the 'characteristic' *K,* a statistic which he shows experimentally to be independent of sample size for homogeneous material. Its chief drawback appears to lie in its sensitivity to variations from work to work in the style of a single author, which in some cases is nearly as great as its sensitivity to the variations among different authors. However, this may merely indicate that, for a given author, the particular quantitative features of vocabulary distribution reflected in values of the characteristic are stable only within groups of works involving very similar ranges of subject matter. As Yule himself has stressed, a great number of control calculations based on new data are necessary before the value and significance of measures such as the characteristic can be established.

Although Yule has been criticized for limiting his study to vocabulary (and to noun occurrences only), he was well aware of these short-

---

[1] Guiraud (1954a) has come out particularly strongly in support of the use of relative frequency criteria in studies of literary style.

comings and modestly presented his work as a far-from-perfect initial exploration of only one of the important aspects of literary style. Rather than being detrimental, however, the limitation of the scope of Yule's investigations contributes to the thorough and intellectually honest treatment of the problems encountered, so that his book as a whole serves as an example of the type of dedicated, responsible scholarship which will be necessary in producing further significant advances in this field.

Portions of recent books by Herdan (1956, 1960) and by Fucks (1955) have been devoted to the treatment of questions relating to the statistics of literary style. In the first work, Herdan defines a statistic v_m, which is very similar to the ‘characteristic’ K, except that it does not depend on Yule’s assumption that word frequencies follow a Poisson distribution; moreover, v_m may be simply described as the coefficient of variation of a mean. Except for this one innovation, for whose significance Herdan makes rather extensive claims on the basis of very limited evidence, his section on stylostatistics is largely a restatement of Yule’s work. In the corresponding section of his second book, however, Herdan proposes a number of original techniques, including the use of area ratios from Lorenz diagrams to measure vocabulary concentration.[1] Unfortunately, extensive control calculations of the sort proposed by Yule have yet to be carried out in connection with these new measures.

Fucks’ work represents a somewhat different approach to literary statistics than that of either Herdan or Yule, since he concentrates on syllables as his fundamental units rather than on words. This leads him to consider certain statistics related to the distribution of the number of syllables per word, as well as others related to the distribution of various metrical and syllable-space patterns. During the course of his treatment of possible metrical statistics, Fucks introduces an intuitively reasonable measure of metrical rigidity which ranges in value from 0 for ‘absolute Prosa’ to 1 for ‘absolut gebundene Rede’. He gives no examples of the application of this simple measure, however, but instead concentrates on the introduction of complex mathematical apparatus, including twelve-dimensional vectors for the representation of the relative frequency of the twelve types of metrical units with which he deals. Although Fucks’ intention is evidently that of outlining a number of possible avenues of approach to the statistical analysis of style, his book nevertheless leaves the

---

1) Herdan (1960), Chapter 2.

impression of being topheavy with mathematical techniques whose applicability in this area remains to be demonstrated. A general n-dimensional framework for the resolution of problems of disputed authorship can hardly be of much practical use until some of the n stylistic measures which it presupposes have been adequately evaluated.

Large areas in the field of literary statistics remain to be explored in the effort to discover new, and hopefully more stable, measures of an author's style than those evolved to date for vocabulary. Investigations on the syntactic level, including the statistical study of patterns of coordination and subordination and of the types and depths of 'nesting' in sentences,[1] may possibly provide interesting new insights into certain aspects of literary style which have yet to be handled quantitatively. An essential prerequisite for the extension of the scope of literary statistics, and for its development into a more powerful analytical tool, is the active participation of linguists in such efforts, preferably linguists proficient in the use of statistical techniques. Literary statistics is a field in which there is no substitute for very careful groundwork, both in the linguistic and in the statistical analysis, before computation is even begun, much less conclusions reached. It is to be hoped that work of the necessary caliber will cease to be the exception in the future and will eventually become the rule through greater cooperation between statisticians and linguists.

INFORMATION-THEORETIC MODELS. Since the appearance of the basic mathematical studies by Wiener (1948) and Shannon (1949), the field of information theory has been the scene of widespread activity[2] which has included the participation of scholars in a number of areas other than that of communication theory. In particular, linguists, psychologists and engineers have made efforts to apply certain information-theoretic concepts and techniques to the study of linguistic problems. Shannon's measure of entropy (or 'selective information') – essentially an average of the statistical 'surprise value' or rarity of signs produced by a source in a given communication system – has received particular attention as a possible vehicle for quantitative statements about language processes and language structure.

As has been emphasized many times in the literature, the measure of information used in information theory has nothing to do with the

---

1) For a treatment of some of these concepts, see the discussion of the work of Yngve and Oettinger in the sections on syntax in Part 3.

2) The reader is referred to Stumpers (1953) for a guide to the pertinent literature.

semantic content of transmitted messages, but deals exclusively with the statistical structure of their formal representations. This immediately rules out the possibility of any application of information theory to the study of semantic questions – a connection one might be tempted to make when confronted with the term 'information' taken out of its precise mathematical context. Cherry[1] has pointed out that analysis of the behavior of a source of signs in terms of Shannon's information measure is valid only when the source is statistically stationary, so that observation of the frequencies of the various signs over a period of time yields accurate estimates of the statistical parameters of the source. In Cherry's opinion, sign behavior is not stationary in most areas of human communication, so that the mathematical theory does not strictly apply. He makes this reservation only with regard to the communicative behavior of individuals, however, leaving open the question of behavioral norms for groups. Herdan,[2] on the other hand, insists that the probability distributions of linguistic units are stable, and that information theory is therefore an appropriate tool for linguistic research.

Whatever the relative merits of the two somewhat conflicting points of view mentioned above, entropy values have been computed for letter and phoneme distributions (which seem to exhibit a reasonable degree of stability) and for the distribution of word length in terms of syllables. In the latter connection, Fucks (1955) has employed Shannon's entropy as a statistic of individual literary style, without drawing upon its role in communication theory. Such use of the selective information measure is, in itself, completely reasonable. It becomes of practical significance, however, only when extensive data have been accumulated on the behavior of the statistic for the class of distributions under consideration.

Of considerably greater interest are those studies which go beyond a purely statistical application of the information measure and include more of the associated theory. One important information-theoretic concept, that of the redundancy, has appeared to be of particular value in modelling certain formal properties of natural language systems in terms of structural characteristics of codes. Briefly, the redundancy of a source (and, by extension, the redundancy of the code employed by the source) is defined as $R = 1—H/H_{max}$, where H is the actual information rate of a source and H_{max} is its maximum information rate, which is realized only if all signs produced by the source occur

---

[1] Cherry (1957), p. 177.

[2] Herdan (1956), Chapter 9.

independently with equal probabilities. This latter condition for $H = H_{max}$, that is, for zero redundancy, implies that there is no hierarchical structure among the signs produced by the source under consideration: there are no preferred sequences or combinations in the code. While making maximum utilization of any given alphabet of signs, a code of this sort has the possibly serious disadvantage that any errors in transmission of a message will of necessity go undetected, since all messages are equiprobable. Natural languages, on the other hand, with their preferred and forbidden combinations of formal units on several structural levels, exhibit a relatively high degree of redundancy, which accounts to a considerable extent for their effectiveness as means of communication under a variety of adverse conditions.[1]

Independent studies by Shannon (1951) and by Miller and Friedman (1957) have employed various psychological tests for the purpose of estimating the entropy and redundancy of printed English texts. Their results indicate that, on the alphabetic level, the redundancy of English is approximately fifty per cent. Miller and Friedman, who tested the ability of subjects to restore printed messages which had been subjected to different degrees and types of mutilation, have also considered the question of how printed English can best be compressed for the purpose of saving channel capacity in transmission. They conclude that the most successful method involves the systematic omission of vowels and the word space – a conclusion of some interest to linguists with respect to both writing systems and processes of phonetic change.

Application of information theory to spoken languages has for the most part been carried out in terms of distinctive feature analysis as developed by Jakobson and his co-workers. Since the distinctive features are presumed to be binary in nature,[2] the use of the ‘bit’ – the binary unit of selective information generally employed by communications engineers – is particularly convenient here. In analyzing the structure of phonemic systems, the number of distinctive features actually in use can be compared with the minimum number of binary distinctions necessary to code each phoneme uniquely.[3] Working along these lines, Greenberg, Osgood and Saporta (1954) have proposed that a simple ratio of these two quantities be considered as a measure of the

---

1) For a fuller discussion of the role of redundancy in natural languages, see Gleason (1955), Chapter 19.

2) Cf. Halle (1957).

3) Cf. Cherry, Halle and Jakobson (1953). Belevitch (1956) also considers this approach.

efficiency of a given phonemic system. A somewhat different approach to the study of distinctive feature analyses will be treated in Part 3 in the section *Set-theoretic Models.*

QUANTITATIVE METHODS IN HISTORICAL AND COMPARATIVE LINGUISTICS. The use of quantitative methods in historical and comparative linguistics is, as has been pointed out by Whatmough (1957), by no means a new phenomenon. Within the field of Indology, for example, a series of studies, beginning with the work of Arnold (1905) on Vedic meter, have incorporated simple frequency criteria as aids in determining the relative antiquity of various portions of the Ṛgveda. During the past several years, however, the relatively limited attention accorded to such earlier attempts has given way to a great upsurge of interest on the part of both linguists and anthropologists in a number of quantitative methods which have been recently proposed for measuring the degree of resemblance among a set of languages. Although a few efforts (some of which will be mentioned later) have been made in the direction of establishing a quantitative basis for a general typology of languages, 'degree of resemblance' has in most cases been carefully qualified in order that it may serve as a criterion for the degree of genetic relationship within a language family. The more ambitious studies of genetic relationship - those employing the techniques of lexicostatistics, or 'glottochronology', as it is sometimes called, have attempted not only to summarize the interrelationships among the members of a given language family in the form of a *Stammbaum,* but to date each of the branching points as well.

The quantitative study of the genetic relationship of languages dates back at least as far as the work of Czekanowski (1927), who attempted to measure the closeness of relationship of pairs of Indo-european languages on the basis of the number of phonological and morphological features (out of a selected list of about twenty features) shared by both members of a pair. Some ten years later, Kroeber and Chrétien (1937) conducted parallel investigations using a more extensive list with seventy-four features. Their basic method entailed the use of correlation formulas defined on four-cell tables, where one of the cells contained the number of features common to a pair of languages, another cell contained the number of features appearing in the first language, but not the second, and so forth. The results obtained by Kroeber and Chrétien using a variety of correlation formulas were very similar, and in general corroborated the prevailing consensus of opinion regarding the interrelationships of the Indoeuropean languages. The authors reported, however, that their findings indicated that the Germanic group was somewhat more closely related to the Balto-Slavonic group than to Italo-Keltic, and that Greek was more closely related to Sanskrit, Armenian and Iranian than to Italic and other *centum* languages.

More recently, Ellegård (1959) has pointed out, as Chrétien (1943) himself did earlier, that a large number of the original Kroeber-Chrétien correlation values are statistically nonsignificant. Ellegård demonstrates that the appearance of the number of features occurring in both languages and the number of features occurring in neither language as symmetric variables in several of the Kroeber-Chrétien formulas is a potential source of seriously misleading results. He proposes a formula which does not involve the latter variable and applies the formula both to Ross' data (Ross, 1950), on common Indoeuropean roots and to the Kroeber-Chrétien data. Despite over-all agreement of his two sets of results, Ellegård finds sufficient discrepancies to conclude that no single statistic can measure the similarity of entire languages as such. He nevertheless envisions the possibility of considerable progress in the measurement of language similarity through the reciprocally beneficial development of a linguistic taxonomy and of statistical methods defined within the taxonomical framework.

Quantitative approaches to the typology of language have been investigated by Menzerath and Meyer-Eppler (1950) and by Greenberg (1954), in both instances on the basis of the formal structure of words. Menzerath and Meyer-Eppler propose the use of three criteria for classifying all words of all languages: the number of syllables per word, the number of phones per word, and the form type of a word in terms of the sequence of vowels and consonant clusters. Languages are then classified according to the distribution of their vocabularies among the various word classes. Greenberg, building on the work of Sapir, develops his classification according to the morphological structure of words. He defines ten indices in terms of the relative frequency of particular types of morphemes as measures of tendencies toward synthesis, agglutination, the use of prefixes, and so forth. While the establishment of a general taxonomy of languages still remains as a major task for the future, it is evident that quantitative methods such as those touched upon here will play a major role in any such development.

The ‘time-depth’ computations of lexicostatistics, as developed by

Swadesh (1950) and elaborated by Lees (1953), are based upon language similarity of a very specific nature: the set of root morphemes corresponding to a carefully assembled list of glosses is constructed for each of a pair of languages, and the number of word pairs which are cognates (as determined by the comparative method) is counted and expressed as a percentage of the total number of pairs examined. The time elapsed between the period from which the language data are taken and the point in the past at which the two languages are assumed to have begun to develop independently is then computed using the formula $i = \log C/2 \log r$. Here i is the elapsed time, or 'time-depth' (usually expressed in millennia), C the percentage of cognate pairs, and r the percentage of pairs retained for unit time.

The derivation (Lees, 1953) and use of the 'time-depth' formula depend on a number of pivotal assumptions, the first being that there exists in every language a set of basic root morphemes so stable that only a small number of them are replaced during a period as long as a thousand years. It is supposed, moreover, that certain of these root morphemes express 'cultural universals' which are common to all languages, so that a list of glosses can be drawn up which will have a (nearly) complete set of corresponding root morphemes in any language; in other words, the list can serve as a tool for finding comparable sets of stable vocabulary items for different languages. The second assumption is that the rate of 'morpheme decay', or gradual replacement of items in the set of basic morphemes, remains essentially constant both with time and from language to language. Finally, when using the formula, it is usually assumed as a first approximation that the languages in question have developed completely independently of one another, although Swadesh (1955) has modified the formula, if only symbolically, to allow for various degrees of mutual influence by introducing the 'separation factor' $\acute{s}$.

The retention rate r in the 'time-depth' formula has been determined from Indo-European data, since the separation in time is relatively accurately known in many instances in this language family. Lees has computed r for thirteen pairs of Indo-European languages using the original Swadesh list of some two hundred glosses and has arrived at an average r of about eighty per cent per millennium. Swadesh (1955), who has applied his formula to problems concerning the genetic relationship of American Indian languages, is somewhat critical of Lees' results and of his own earlier efforts. In particular, he cites the need for a more general control study than that of Lees for establishing the rate of retention and proposes further study of the test list of glosses, which he himself has improved and reduced to one hundred items. Despite its current imperfections, however, Swadesh expresses great hopes for the future development of lexicostatistics, which he ' envisions as potentially becoming a 'precision instrument' in historical-comparative studies.

Criticism of the lexicostatistical method as presented by Swadesh and Lees has come from several quarters. Hoijer (1956) has made statements to the effect that no test list can be devised which will work for all languages, since even when ‘cultural universals’ are involved, there may be no unique one-word correspondent in a given language. He cites a number of examples from Nahavo to support this contention, maintaining that in each case where multiple correspondents exist one might, depending on one’s choice, find either retention or loss of a vocabulary item and thereby obtain results which are seriously in error. Kroeber (1955) is particularly critical of the use of computed ‘time-depths’ as evidence of relationship when the percentage of shared vocabulary is less than ten per cent, since borrowing or chance similarities in just a handful of words would greatly alter the results in such cases. For this reason, he suggests that lexicostatistics be applied primarily to the study of sure genetic groups, rather than to the detection of distant relationships.

One of the most interesting studies, both critical and exploratory, to appear on lexicostatistics is the work of Arndt (1959) on Germanic. Arndt’s underlying aim is twofold: to test the validity of the glottochronological method and, in turn, to use the method to test the leading non-statistical theories regarding the historical relationships obtaining among the various Germanic languages. His results concerning the second problem point very strongly to the rejection of the tripartite theory of quasigenetic descent and give some support to a grouping into East Germanic, North Germanic, North-sea Germanic and Inland Germanic.

Despite the apparent usefulness of lexicostatistical methods in casting light on such theoretical questions, Arndt finds evidence that they suffer from several shortcomings. One of his major points is that application of the Swadesh formula to pairs of modern dialects consistently gives much ‘shallower’ time-depths than it does for the corresponding pairs of old dialects, and, moreover, that it is precisely for the ‘old-old’ pairings that the most historically plausible dates of separation are obtained. He concludes that this is evidence of strong mutual influence and synthesis, and that any averaging together of the divergent results must be rejected on those grounds. (Similar observations have been made by Rea (1958), who applied the lexicostatistic dating method to Romance languages and obtained divergence dates ranging from 800 A.D. to 1600 A.D., instead of the actual figure of somewhere around 100 A.D.) Arndt terminates his discussion by raising a number of questions regarding the use of the test list, possible short-range fluctuations in the retention rate and the possible effects of the rising literacy rate on vocabulary stability. The existence of so many sources of potential inaccuracy leads him to express grave misgivings as to the value of glottochronology in those linguistic fields for which the available historical and linguistic data are much more limited than in Indo-European studies.

Gleason (1959) has proposed that the concept of lexico-statistics be broadened to include non-glottochronological techniques for determining the most probable genetic relationship among the members of a group of languages. The two methods which he suggests avoid many of the more controversial assumptions mentioned above, since they involve neither dating nor the necessary use of gloss lists designed to include only universal concepts. The first method, that of 'counter-indications', makes use of a count of the number of cases where the words corresponding to a given gloss are not cognates in a given pair of languages, but where each word is a cognate shared with at least one other language in the group. The basic assumption involved in using this approach is that any discontinuity in behavior (such as the switch from one word to another for expressing the same concept) is exceptional, so that the Stammbaum configuration which minimizes the counter-indications is chosen as the most probable.

Gleason’s second method, that involving the ‘characteristic vocabulary index’, uses a two-way table in which the names of the languages in the group under study appear as headings of both the columns and the rows. Each set of cognates found for an item in the gloss list is assigned the value unity. If three languages share a cognate, the number 1/3 is added in each of the cells determined by the intersection of the appropriate rows and columns; if seven languages share a cognate, then 1/7 is added, and so forth. When the process has been completed, the relative size of the totals in the various cells is interpreted as a measure of the closeness of relationship between the corresponding pairs of languages.

In addition to their evident intrinsic appeal as straight-forward methods which can complement and check results obtained using the glottochronological approach, Gleason’s techniques have the added attractive feature that they are designed in such a way that clerical help untrained in linguistics can perform the bulk of the counting and calculating work by following an automatic procedure suitable for conversion into a digital computer program. Thus the entire process can be greatly speeded up, sparing the linguist most of the drudgery usually associated with such work, and allowing him to concentrate on the later stages of the process, which require his linguistic judgement and intuition.

It is indeed fortunate that the widespread activity current in lexicostatistics and related fields has attracted not only able innovators, but keen critics as well. The use of computers to carry the heavy end of the associated data processing load should have two highly desirable effects: On the one hand, it should free linguists for a more concerted attack on the purely linguistic problems involved, and on the other hand it should make the prospect of performing extensive control calculations – an essential pre-requisite for the accurate evaluation of results – seem considerably less formidable than it has in the past.

3. STRUCTURAL MODELS OF LANGUAGE

SET-THEORETICAL MODELS. In the course of the past several years, various structural models have been proposed for the representation and analysis of a wide range of linguistic phenomena. A brief perusal of the related literature makes it quite evident that many of the scholars contributing to this trend have drawn heavily on the fields of modern algebra and symbolic logic. Not only have numerous concepts and representational methods been borrowed, but, in a number of instances, the entire metalinguistic presentation has faithfully followed the pattern of deductive mathematical systems.[1] A consideration of several studies involving the application of set theory and related logical concepts to certain aspects of natural languages should help to illustrate the nature of current developments in this area.

In a recent article, Harary and Paper (1957) describe the application of several concepts from relation theory to the development of a rigorous mathematical framework for the description of phonemic distribution. Starting with the set P of all phonemes in a language and the relation R 'is immediately followed by', the authors proceed to

1) Some scholars tend to emphasize this area of mathematical linguistics to the exclusion of those in which mathematical methods play a more secondary role. Cf.. for example, Koschmieder (1956).

describe the distribution of a given phoneme in terms of six familiar properties of binary relations, including symmetry, reflexivity, and transitivity. A typical distributional statement in their system appears in the form 'phoneme *x* is symmetric with respect to phoneme *y*' if the sequences /*xy*/ and /*yx*/ both occur in the language under consideration. A quantitative level is then superimposed on the basic algebraic structure of the model through the definition of a set of simple numerical indices whose values are determined by the fraction of cases in which the distribution of a phoneme exhibits a given relational property.[1] The indices introduced here are very reminiscent of those suggested by Greenberg (1954) for use in typological studies on the morphological level, and the authors themselves have proposed that their model be applied to the typological investigation of phonemic distribution. Whatever the degree of success of future work in such directions may be, the authors' important claim that their model makes concepts such as 'freedom of distribution' more precise appears to be well supported by their examples. It is questionable, however, just how far the proposed extension of their model to n-place sequences can proceed without sacrificing the clarity and simplicity of the current version.

A number of algebraic and logical representations of phonemic systems are introduced by Ungeheuer (1959a) in his investigation into the logical foundations of Jakobson's distinctive feature theory. Ungeheuer first treats the different features as classes with defined complements, representing them alternatively as binary variables entering to Boolean functions.[2] With the aid of these models, he is able to give a clear presentation of such important points as the distinction between the 'natural' redundancy of a sound system and that of the formal system resulting after analysis, which usually declares certain features irrelevant. In Ungeheuer's second approach, motivated by certain psychological considerations, only the presence of the positive value of a feature is regarded as significant. Proceeding on this basis, he describes the 'space' of all possible phonemes in set-theoretic terms as the set N of all subsets of the set of distinctive features. As he points out, N may be treated mathematically in several ways, for example as a Boolean algebra with the features as generators. But

---

[1] This is an illustration of the fact that the distinction between quantitative and structural models is by no means a strictly dichotomous one; a given model can exhibit both structural and quantitative characteristics.

[2] Belevitch (1956) has also mentioned the use of Boolean functions in this connection.

Ungeheuer sounds an appropriate note of caution with respect to reading too much into such models: any application of the highly developed body of mathematical theory associated with them must await fuller understanding of the underlying phenomena.

The construction of machine translation grammars based on set-theoretic concepts is the subject of a recent study by Kulagina (1958). Undefined concepts such as those of ‘word’ and ‘grammatical phrase’ play the role of axioms in the formal development of Kulagina’s system. Upon this foundation she erects a hierarchy of grammatical classes by means of the rigorous definition of a series of partitions of the set of words into disjoint subsets. On the lowest level, that of the partition into word ‘families’, the members of a given subset may be substituted for one another in any grammatical phrase without making it ungrammatical, and conversely. Proceeding to successively higher levels, Kulagina introduces enough formal apparatus to be able to handle some of the simpler concepts of French grammar; she explains that Russian, on the other hand, will not fit into her system, since for any given grammatical case there are three ‘families’ of nouns in the singular, and only one in the plural. The proposed model has certain other limitations as well, one of them stemming from the fact that the possibility of homographic usage of a single form as more than one part of speech – a problem which cannot be ignored in machine translation – is not provided for in a system of mutually exclusive grammatical categories.

Before turning to the consideration of syntactic models, it will perhaps be appropriate to mention the serious efforts on the part of linguists of the Glossematic School to develop a ‘calculus of non-quantitative functions’ – a new algebra aimed primarily at the description of humanistic materials such as language. Glossematic algebra, as presented by Uldall (Hjelmslev and Uldall, 1957), employs much of the terminology and notation of logical set theory, but with the important difference that certain familiar concepts, such as those of function and negation, are defined in an unusual manner. At least one critic (Ungeheuer, 1959b), moreover, has pointed out that glossematic terminology is not always used consistently by Hjelmslev and Uldall, apparently due to the confusion of concepts carefully distinguished in formal logic. On the basis of what the glossematicians have presented so far, it appears that the development of their algebra would have profited greatly from the use of an established logical framework.

MODELS FOR SENTENCE SYNTHESIS AND SYNTACTIC DESCRIPTION. The remainder of Part 3 will be devoted to the consideration of syntactic models. The first subsection, which deals with syntactic description and sentence synthesis, will be followed by a section on models designed for use in syntactic analysis. Although the studies to be treated in the latter section vary considerably in scope and emphasis, they all share the common aim of providing systematic mechanical procedures for the syntactic analysis of textual material, primarily in connection with automatic language translation. The present section represents a somewhat less homogeneous grouping: because of the body of theory common to them all, Yngve's model for sentence generation (Yngve, 1960) will be treated together with primarily descriptive models such as those of Harris (1957) and Chomsky (1957). Chomsky's threefold classification of grammars according to whether they are represented as finite state, phrase structure, or transformational devices will be adopted as a convenient framework for the discussion which follows, although the emphasis will be somewhat different from that in his book.

A finite state grammar may be characterized (Chomsky, 1957) as a machine with a finite number of internal states which generates1 the sentences of a language in the following manner: starting in a unique *initial state,* the machine passes into a second state by producing the first word of a sentence, then, while emitting a word at each transition, proceeds from state to state until it has reached the *final state,* at which time a complete sentence has been generated. Such a machine is a special case of a finite-state Markov process, and hence can be represented by a graphical device often used in connection with such processes: a 'state diagram', which represents each state by a point or small circle and each allowable transition by an arrow connecting the appropriate points. This type of diagram, generalized to include an assigned probability for each transition, is often employed in communication theory to model the statistical properties of a source which has a finite alphabet of signs.[2]

1) Chomsky (1959) has modelled the properties of various types of grammars in terms of automata, by associating with each type an automaton powerful enough to generate sentences according to the rules of its companion grammar. Lees (1959) has remarked that the grammars developed by Chomsky "generate" the sentences of a language only in the very specialized sense that they provide precise rules for constructing all sentences of the language. Such grammars are designed primarily to provide a compact description of syntax, rather than for the generation of sentences in a computer-programmed system.

2) Cf. Shannon and Weaver (1949), pp. 15 ff. and also the discussion under *Information-theoretic Models* in Part 2 of this paper.

In one of his earlier articles on machine translation, Yngve (1955) proposes the use of finite state representations as a possible means for compressing a grammar powerful enough to handle entire sentences into a compact form which could be stored in a computer. As a first step in achieving this compression, each state-to-state transition is to have associated with it a part of speech, rather than a single word. Further simplification is to be attained by grouping paths together into phrase types, and these, in turn, into clause and, finally, sentence types. The latter stages of the proposed multiple-level finite state grammar are not worked out in any detail, so that it is difficult to judge its potentialities directly. Chomsky (1956, 1957), however, has shown that finite state grammars are inherently incapable of representing the recursive properties of certain English constructions, where recursion is allowed to proceed without limit. Although such theoretical arguments cannot rule out the possibility of constructing a finite state grammar capable of handling sufficiently involved sentences for practical machine translation work, they give a strong indication that such a grammar would probably be prohibitively complex.

The second grammatical model which Chomsky considers, that of phrase structure, or immediate constituent analysis, has been in use among American linguists for some time in one form or another.[1] Chomsky has formalized the theory of phrase structure grammar by treating it as an operational system (Σ, F) consisting of a finite set Σ of initial strings, together with a finite set F of phrase structure rules of the simple form $X \rightarrow Y$. The rules of F operate one at a time on strings to produce new strings, each rule specifying the replacement of a single symbol by one or more other symbols.

In all cases of interest, repeated application of the phrase structure rules, beginning with one of the initial strings, produces a sequence of strings which terminates in a string that cannot be further modified by the rules of F. Such unmodifiable strings, called *terminal strings,* are the sentences of the language described by the (Σ, F) grammar. The sequence of strings produced during the generation of any given sentence determines its constituent structure, which may conveniently be represented in a tree-like branching diagram.

Phrase structure grammars, as shown by Chomsky, do not suffer from the same failings as finite state grammars, since they are capable of mirroring the recursive properties Of natural languages through the inclusion of recursive rules in F. He has emphasized, however, that

---

[1] The work of Wells (1947), who attempted to unify and systematize the results of previous efforts in this area, is particularly well-known.

there are types of English constructions – in particular, those involving discontinuous constituents – for which the phrase structure model offers what is at best an extremely awkward description. Rather than suggesting that the model be elaborated on the level of phrase structure description (a course which he regards as unsound).[1] Chomsky has proposed the addition of a higher structural level, that of grammatical transformations, as a remedy for such difficulties. Before embarking upon a discussion of transformations, however, it seems appropriate to consider a model of automatic sentence synthesis which incorporates much of the apparatus of simple phrase structure

Some of Yngve's more recent work has involved the investigation of models of sentence construction which are suitable for mechanization as part of the output stage of an automatic translation system. Because of this machine-oriented approach, Yngve's basic model (Yngve, 1960) incorporates the necessary restriction that its sentence-generating mechanism have finite temporary storage. Using phrase structure rules of the type described above,[2] the mechanism generates sentences from left to right, the left-most symbol of a given string always being expanded before the symbol immediately to its right, and so forth, until a terminal string is produced. Intermediate results are stored in the temporary memory, a linear storage array with only one and accessible for read-in or read-out of information, in accordance with the principle of exclusively left-to-right expansion of strings. This type of storage array, which has been referred to in the literature as a 'pushdown store' (Oettinger, 1960b) has been shown to have important applications to automatic syntactic analysis, some of which will be dealt with below.[3]

One interesting aspect of the output of Yngve's model is that an explicit representation of the phrase structure of each sentence is included along with the words out of which it is formed. This representation takes the form of the Łukasiewicz (1957) parenthesis-free notation, each node of the phrase structure tree being followed immediately by the subtrees which branch out from it. In connection with an investigation of the adequacy of his model, Yngve has shown that there is a direct relationship between the tree structure of a sent-

1) Chomsky (1957), footnote 6, pp. 41-2.

2) Yngve has also proposed a conceptually simple scheme for incorporating rules of the type $A \rightarrow B + \cdots + C$ into the system, so that it can handle certain discontinuous constituents.

3) Cf. the discussion of predictive analysis in the final section of Part 3.

ence and the maximum amount of temporary storage required in the course of its generation. On the basis of a right-to-left numbering of the branches originating at each node of a tree, he defines the 'depth' of each terminal node as the sum of the numbers written on all branches leading to it. The maximum terminal node depth of the entire tree, referred to as the depth of the sentence, turns out to be the number of temporary storage locations needed to produce the sentence.

Yngve has shown that the depth of a given sentence is strongly dependent on the direction of the various branches of its associated tree. Sentences whose trees branch primarily to the left have relatively large depths; those with primarily right-branching structure have relatively small depths. The two structural types are referred to by Yngve as *regressive* and *progressive,* respectively. From an investigation of depth phenomena in English sentences, Yngve has concluded that there is an upper bound on the depth of English sentences – a bound which agrees rather closely with a figure given by Miller (1956) as the approximate span of immediate memory for human subjects. Yngve has noted, furthermore, that English contains a variety of grammatical devices, such as discontinuous constructions and the introductory use of 'it', which serve to decrease the depth of sentences while maintaining the expressive power of the language. Much of the apparently unnecessary complexity of English grammar can be accounted for, he feels, in terms of the use of alternative constructions for the purpose of reducing sentence depth.

Yngve has generalized his findings for English in the form of an hypothesis which predicts that the behavior of all languages will be found to be essentially similar with regard to depth phenomena. The hypothesis clearly has important implications not only for linguistic theory, but for psychological theory as well. If the hypothesis can be verified for other languages, there will be good reason to adopt Yngve's sentence generation mechanism as a model of human language production, whatever its ultimate value in machine translation work turns out to be.

Returning now to the subject of grammatical transformations, a concept first formalized by Harris (1952) in connection with his work on discourse analysis, we shall consider the recent treatment of transformational analysis, as presented in the work of both Harris (1957) and Chomsky (1956, 1957). Although there has been a considerable amount of idea-sharing between Harris and Chomsky, as both authors take pains to point out, their work has proceeded along fairly independent lines. Harris has framed his development of grammatical transformations in terms of the co-occurrence of morpheme classes in the sentences of discourse. He introduces transformations as a means of relating different constructions which have the same (or nearly the same) co-occurrence classes. Once this additional level of transformations is established, Harris claims, there is no longer any need to analyze all sentence types independently: those with the same co-occurrence clauses as previously analyzed sentences can be treated as transforms of the latter. Moreover, complex sentences and certain sentence sequences can often profitably be regarded as the concatenation of one sentence with the transform of another.

In the more formal portions of their respective developments of grammatical transform theory, both Harris and Chomsky accord a central position to what they term the *kernel sentences* of a language. These are the elementary underlying sentences from which all other sentences of the language are obtained by means of transformations.

In Chomsky's model of transformational grammar, the kernel contains precisely those simple, active, declarative sentences which can be obtained by use of elementary phrase structure rules. The transformational portion of the grammar, which constitutes a level independent of the phrase structure part, contains rules for converting strings with a specified constituent structure into new strings with a different constituent structure. These transformation rules, when applied to the simple sentences of the kernel, are capable of generating all the remaining sentences of the language. Chomsky has emphasized that transformational grammars are inherently more powerful than phrase structure grammars, not only because of the simplifications attendant on the use of two separate levels for syntax,[1] but also as the consequence of a fundamental difference in the nature of the two sets of rules. Whereas the symbols of a given string contain all the information needed to determine what phrase structure rules may be applied to that string, the applicability of a given transformation depends in general on a knowledge of previous steps in the derivation of the string; for example, on whether or not a given substring is an expansion of the symbol NP and may legitimately undergo transformations applicable to noun phrases[2].

---

1) In addition to the two levels for syntax, Chomsky's transformational model has a third level consisting of morphophonemic rules for converting strings of words into strings of phonemes.

2) The output of Yngve's mechanism for sentence generation contains precisely such information, and it would therefore serve as a suitable input to a hypothetical

According to Chomsky, transformational grammars, besides providing simpler and more powerful techniques for describing the syntax of natural languages, have additional desirable properties. Among these additional features is the fact that the ambiguity of such constructions as “the shooting of hunters” can be conveniently explained as follows in terms of the transformational model: the cited construction is a transform of two different kernel sentences, one with ‘hunters’ as subject, the other with ‘hunters’ as object. Worth (1958) has recently attempted to make use of this latter property of transformational analysis as a basis for classifying Russian instrumental constructions. In the conclusion to his study, Worth comments very favorably on the usefulness of transformational techniques, finding that they not only make possible more refined groupings of constructions, but provide a convenient framework for statements on the obligatoriness of modifiers as well. Due to its limited scope, Worth’s study can by no means be interpreted as a demonstration of the practical feasibility of constructing complete transformational grammars for natural languages. Its success would seem to indicate, however, that the time for full-scale testing of Chomsky's model is now at hand. As one of his foremost supporters has indicated (Lees, 1957), validation of Chomsky’s entire theory of grammar will ultimately depend on the results of applying it to a wide range of languages.

MODELS FOR SYNTACTIC ANALYSIS. It is a well-known fact that the artificially constructed notational systems of logic and mathematics exhibit a regularity and simplicity of syntactic pattern not found in the case of natural languages. As Yngve (1960) has noted, the syntactic system of a given natural language may contain a variety of alternative devices for expressing a single grammatical relation such as that of subordination, whereas certain formal logical languages characteristically employ one uniform method of expression in such cases. Despite the considerable difference in their respective degrees of complexity, however, sufficient over-all similarities obtain between the syntactic systems of natural languages and those of logical languages to ensure that the latter can model important properties of the former in a nontrivial manner. This close relationship is proving to be of particularly great significance in the development of automatic methods of syntactic analysis in connection with machine translation.

A central concept in the theory of syntax of formal logical languages

mechanism for performing the operations of the transformational part of Chomsky’s grammar.

-- that of well-formation of a formula – corresponds very closely to the concept of grammaticalness as conventionally applied to the phrases and sentences of natural languages. Well-formed formulas, like grammatical sentences, are those which have been constructed according to the rules; in both cases, it is of obviously great practical importance that there be some reliable means for distinguishing well-formed sequences from non-well-formed or ungrammatical sequences, for otherwise the language in question loses its value as a means of communicating information. In automatic translation work, a test for well-formation is of prime importance, both as a means for determining the boundaries of well-formed subconstructions, and as a valuable check on the validity of the over-all analysis of entire sentences. For reasons such as these, some of the earlier theoretical studies connected with machine translation were devoted to the exploratory development of syntactic calculi which might later be employed as the basis of machine routines for testing the grammaticalness of sentences.

The work of Bar-Hillel (1953), based in part on some of the ideas of the logician Ajdukiewicz (1935), is fairly characteristic of a number of the efforts in this direction. Starting with the two primitive symbols *s* and *n* (for sentence and noun, respectively), Bar-Hillel has attempted to construct a system of grammatical notation which would make possible the use of a particularly simple test for well-formation, patterned after the arithmetic process of cancellation of fractions. Verbs, which combine with nouns to form sentences, are assigned the symbol *s/n,* for example: a sequence of a noun followed by a verb will automatically be analyzed as a sentence, since the *n* for the noun will be cancelled by the *n* in the 'denominator' of the symbol for the verb, leaving only the symbol *s.* Bar-Hillel defines a string of his grammatical symbols as 'syntactically connex' if there is some sequence of cancellation operations which can reduce the entire string to a single primitive symbol. The syntactically connex strings represent either entire grammatical sentences or well-formed constituents of sentences.

As Bar-Hillel himself has made clear, his model suffers from several inadequacies, among them the fact that the original string of symbols assigned to a given sentence can be expressed in a number of ways, depending upon one's view of how the elements of the sentence combine. Further difficulties, similar to those encountered in phrase structure grammar with regard to the treatment of discontinuous constituents, are caused by the fact that only immediately adjacent symbols may cancel one another. Lambek (1958), who has developed a model of syntax similar to Bar-Hillel's, has proposed the following method for circumventing both of these difficulties: the words of the sentence under consideration are first bracketed in all possible ways; next, all possible grammatical types are assigned to each word. The type of each of the resulting expressions is then computed in accordance with the grouping of elements indicated by the parentheses. Although the existence of such a finite exhaustive procedure may satisfy a pure mathematician or logician that the problem is solved, 'solutions' of this sort are unfortunately far too complex and general in nature to lead to economically feasible machine translation procedures.

In contrast to theoretical studies of syntactic calculi, papers describing practical efforts to solve some of the problems of automatic syntactic analysis have only recently begun to appear in the literature. Several machine translation groups have reported a fair degree of success in the automatic recognition of minor constructions, such as prepositional phrases and noun phrases, usually on the basis of a machine search of a limited context. There are two main difficulties attendant on such 'microsyntactic' methods of syntactic analysis, however; the first is that the indicated constructional groupings are sometimes incorrect because of the decisive influence of higher-order constraints, whose nature cannot be determined on the basis of a simple scan alone. The second difficulty is that there seems to be no clear way of extending the methods of microsyntactic analysis to provide a viable approach to the analysis of clause and sentence structure; neither extension of the context scanned nor iteration of the procedure appears to offer a solution.

A number of workers in the machine translation field, recognizing that the microsyntactic approach in a sense leads up a blind alley, have been looking for more powerful techniques capable of dealing with the syntactic analysis of entire sentences. Some have envisioned a two-level scheme of analysis, wherein a 'macrosyntactic' routine (applied to the output of the microsyntactic routine) determines the large-scale structural features of each sentence, while correcting any errors made on the lower level of analysis. More desirable from the point of view of both elegance and cost, however, would be a scheme which could handle the entire process of automatic syntactic analysis in a single uniform framework. The system of predictive syntactic analysis originally proposed by Rhodes (1959a, b,) of the National Bureau of Standards and elaborated and modified by Sherry (1960), Oettinger (1960b), and others[1] at Harvard appears not only to fulfill

---

1) Cf. Bossert (1960) and Bossert, Giuliano, and Grant (1960).

this latter requirement, but to have other characteristics which are highly desirable in machine translation work.

Since a detailed account of the mechanics of predictive analysis is available in the literature just cited, only a few of the more salient characteristics of the process will be mentioned here. Perhaps the most important consideration from the point of view of machine operations is that in predictive analysis the items composing each sentence are processed one at a time in order during a single left-to-tight scan. Each word in the sentence is tested to determine whether or not it 'fulfills' any of a list of 'predictions' located in a linear array Of storage locations known as the prediction pool. The predictions are essentially statements on the nature of the different syntactic units, On all structural levels, which are expected to occur in a given sentence. Certain predictions, such as those for subject and predicate, are included in the prediction pool at the beginning of the analysis of each sentence; the rest are determined in the course of analysis according to the syntactic role assigned to words previously encountered. The prediction pool operates on the 'last-in-first-out' principle of a pushdown store, so that the predictions added by the previous word will always be the first to be tested for fulfillment by the current word. A given word is assigned the syntactic role specified in the first prediction that it can fulfill; for example, a finite verb will fulfill a prediction of 'predicate head', and will be assigned 'predicate head' as its syntactic role; this will be followed by the removal from the prediction pool of the predicate head prediction and the subsequent addition to the pool of a prediction for an object in the case governed by the verb. The 'pushdown' characteristics of the prediction pool have been found to offer particular advantages in the analysis of 'nested' constructions with discontinuous constituents.

The system has two important self-checking features: the first is that a record is kept on a secondary output of all instances in which any word in a sentence satisfies more than one of the predictions in the prediction pool. Since only one syntactic role is assigned to each word, there is always the possibility in such cases that an incorrect choice was made. The presence of the secondary output ensures the availability for future review of all the information obtained during the analysis, should an inconsistency be discovered at a later point. The second feature is that failure of a given word to satisfy any of the predictions in the prediction pool is interpreted as an indication that there is a break in the 'syntactical connexity' at that point, either as a result of incorrect analysis, or as a consequence of actual ill-formation, of the sort which can be caused by the presence of typographical errors.

The experimental results achieved so far by Oettinger and his co-workers in connection with automatic syntactic analysis of both Russian and English indicate that predictive analysis, while by no means a panacea for all the problems of machine translation, shows considerable promise as a general analytical tool for dealing with the syntactic systems of a variety of natural languages. The work on Russian syntax has progressed to the point where the general run of sentences in scientific texts are likely to be correctly analyzed on a single pass. A further important feature of the system is that despite the simultaneous handling of various syntactic levels, errors in analysis on the higher levels, whatever their cause may be, in general do not prevent correct analysis of lower-order constructions. Thus if a sentence as a whole is not analyzed properly, well-formed subpieces, such as prepositional phrases, are often correctly identified.

This feature of the predictive technique may prove to be of great value in the design of automatic programming systems, where translation is between various symbolic programming languages and the instruction code language of the machine. Rather than stopping the compilation process with the detection of the first error, as is the common practice with systems currently in use, a system incorporating predictive techniques could profitably be allowed to proceed, since it could be counted on that individual subparts of the over-all program would come through the process correctly checked. Oettinger has formalized the statement of this property in terms of what he calls the 'Δ_m-theorem', which he has proved in full rigor (Oettinger, 1960b) for a number of fail-safe algorithms for predictive translation between various formal languages.[1]

A recent paper by Sherry and Oettinger (1960) has still further exploited the relationship of natural and artificial languages; in this case through the construction of a new formal language for the express purpose of modelling the behavior of natural languages when subjected to predictive syntactic analysis. Through the definition of a series of abstract models, each of which incorporates more features of natural languages than does its predecessor, the authors give a step-by-step description of predictive analysis which points clearly not only

---

[1] Wundheiler and Wundheiler (1955) stated at that time that it is possible to translate between various 'frontal' languages (such as that of the Łukasiewicz parenthesis-free notation) with the aid of a dictionary alone. This has been verified experimentally in Oettinger's work.

to its present capabilities, but to areas in which the method must be Considerably improved in the future.

4. APPLIED MATHEMATICAL LINGUISTICS

As has been noted at several points in the course of the preceding discussion, certain of the models and techniques of mathematical linguistics have important practical applications outside the immediate sphere of linguistics itself. Statistical methods of language analysis, for example, have long played an important role in the companion fields of cryptography and cryptanalysis. Due to the well-known conditipns of secrecy surrounding advances in these applied disciplines, however, there has of course been little opportunity for linguistics to draw reciprocal benefit from them. This situation is probably particularly unfortunate with regard to efforts on the part of linguists to decipher ancient written records, an area in which the experience accumulated in scientific code-breaking work might be expected to aid further progress significantly.

During the past decade, the commercial availability of large-scale automatic data-processing systems and the pressure of an enormously increased rate of publication, particularly in the realm of technical literature, have provided dual stimuli to the development of automatic systems for information analysis and retrieval,[1] to be used in the libraries and offices of government, industry, and private institutions. Many of the techniques of mathematical linguistics, including not only those associated with the construction of model languages, but some of a statistical nature as well, are making an increasingly important contribution to this effort. There is a striking kinship, for example, between certain problems connected with automatic literature abstracting, as treated by Luhn (1958) and others, and those encountered in statistical studies of literary vocabulary.[2] In view of such close relationships, future developments in the field of information retrieval should certainly be of more than passing interest to linguists.

The field in which techniques of mathematical linguistics are currently being applied most extensively, however, is undoubtedly that of automatic language translation. Although no attempt will be made

---

1) The reader's attention is called particularly to the publications of the U.S. Patent Office (1957-) and to the papers published in *Preprints of Papers for the International Conference* on *Scientific Information* (1958).

2) Cf. *Statistics of Style and Authorship* in Part 2 of this paper.

here to review the large body of literature[1] which has appeared on the subject, the numerous references made to machine translation work in Part 3 of this paper should give some indication of the tremendous impetus which has been imparted to mathematical linguistics from that quarter. Beyond the potential contributions to linguistic theory which may be by-products of such developments as predictive translation and Yngve's 'depth hypothesis',[2] the present activity in the field of machine translation should have other far-reaching effects on the future course of linguistics as a whole. The extensive experience being accumulated by linguists and mathematicians in the application of large-scale automatic data processing techniques to the material of natural languages should provide a valuable practical groundwork for future linguistic research in areas other than that of translation. As Sgall (1959) has pointed out, advances in modern computer technology have made it possible for the first time to 'fit' a language system into a machine, so that statistical and structural studies of language can now be contemplated on a scale which would have been unthinkable only a decade ago.

In view of the key role which computer technology plays in several aspects of their science, mathematical linguists should take particular satisfaction from observing the current application of some of their own techniques to the synthesis of automatic coding systems for digital computers. As has been remarked earlier, efforts to translate among natural languages by machine have in one instance led to the discovery of simple fail-safe algorithms for translating among formal logical languages, a development with important implications for the future evolution of automatic machine programming. This chain of events might be regarded by some as a rather unusual case of poetic justice, but it is really nothing more than a particularly clear example of the close and mutually beneficial relationship which is being formed between the disciplines of mathematics and linguistics.

WARREN PLATH

---

1) For a guide to the literature on machine translation, the reader is referred to Delavenay and Delavenay (1959) and to the journal *Mechanical Translation* (1954-).

2) Cf. Part 3 of this paper.

BIBLIOGRAPHY

AJDUKIEWICZ, A. (1935), "Die syntaktische Konnexität", *Studia Philosophica* (Łwow), Vol. 1, pp. 1-27.

ARNDT, W.W. (1959), "The Performance of Glottochronology in Germanic", *Language,* Vol. 35, pp. 180-92.

ARNOLD, E.V. (1905), *Vedic Metre in its Historical Development* (Cambridge).

BAR-HILLEL, Y. (1953), "A Quasi-Arithmetical Notation for Syntactic Description", *Language,* Vol. 29, pp. 47-58.

BELEVITCH, V. (1956), *Langage des Machines et Langage Humain* (Bruxelles, Office de Publicite).

BELEVITCH, V. (1959), "On the statistical laws of linguistic distributions", *Annales de la Societé Scientifique de Bruxelles,* Vol. 7B, pp. 310-26.

BOOTH, A. D., BRANDWOOD, L., and CLEAVE, J. P. (1958), *Mechanical Resolution of Linguistic Problems* (New York, Academic Press; London, Butterworths).

BOSSERT, W. (1960), "The Implementation of Predictive Analysis", *Mathematical Linguistics and Automatic Translation,* Rpt. No. NSF-4, Harvard Comp. Lab., Sec. VIII.

BOSSERT, W., GIULIANO, V. E., and GRANT, S. B. (1960), "Automatic Syntactic Analysis of English", *Mathematical Linguistics and Automatic Translation,* Rpt. No. NSF-4, Harvard Comp. Lab., Sec. VII.

CHERRY, E.C. (1957), *On Human Communication* (New York, Wiley).

CHERRY, E. C., HALLE, M., and JAKOBSON, R. (1953), "Toward the Logical Description of Languages in Their Phonemic Aspect", *Language,* Vol. 29, pp. 34-46.

CHOMSKY, N. (1956), "Three Models for the Description of Language", *IRE Trans. on Information Theory,* Vol. IT-2, pp. 113-24.

CHOMSKY, N. (1957), *Syntactic Structures* ('s-Gravenhage, Mouton).

CHOMSKY, N. (1959), "On certain Formal Properties of Grammars", *Information and Control,* Vol. 2, pp. 137-67.

CHOMSKY, N. and MILLER, G. A. (1958), "Finite State Languages", *Information and Control,* Vol. 1, pp. 91-112.

CHRÉTIEN, C. D. (1943), *U.C.P.L.,* Vol. 1, pp. 11-20.

CZEKANOWSKI, J. (1927), *Wstep do historii Slowian* (Łwow; Poznan, 1957).

DELAVENAY, E. and DELAVENAY, K. (1959), *Bibliography of Mechanical Translation* (International Study Group on Mechanical Translation, UNESCO).

ELLEGÅRD, A. (1959), "Statistical Measurement of Linguistic Relationship", *Language,* Vol. 35, pp. 131-56.

FUCKS, W. (1955), "Mathematische Analyse von Sprachelementen, Sprachstil und Sprachen", *Arbeitsgemeinschaft für Forschung des Landes Nordrhein-Westfalen* (Köln und Opladen, Westdeutscher Verlag).

GLEASON, H. A. (1955), *An Introduction to Descriptive Linguistics* (New York, Holt).

GLEASON, H. A. (1959), "Counting and Calculating for Historical Reconstruction", *Anthropological Linguistics,* Vol. 1, No. 2, pp. 22-33.

GLEASON, H. A. (1960), "Genetic Relationship in Language", *Proceedings of the Symposia in Applied Mathematics,* Vol. 12 (Providence, American Mathematical Society) (in press).

GREENBERG, J. H. (1954), "A Quantitative Approach to the Morphological Typology of Language", in R. Spencer (ed.), *Method and Perspective in Anthropology, Papers in Honor of Wilson D. Wallis* (Minneapolis), pp. 192-220.

GREENBERG, J. H., OSGOOD, C. E., AND SAPORTA, S. (1954), "Language Change", Sec. 6.3 of Osgood, C. E. and Sebeok, T. A. (eds.), *Psycholinguistics: A Survey of Theory and Research Problems,* Supplement to *International Journal of American Linguistics,* Vol. 20.

GUDSCHINSKY, S. C. (1956), "The ABC's of Lexicostatistics (Glottochronology)", *Word,* Vol. 12, pp. 175-210.

GUIRAUD, P. (19543), *Les Caractères Statistiques du Vocabulaire* (Paris, Presses Universtaires de France).

GUIRAUD, p. (1954b), *Bibliographie Critique de la Statistique Linguistique* (Utrecht, Spectrum). Revised and completed by T. D. Houchin, J. Puhvel, and C. W. Watkins, under the direction of J. Whatmough.

HALLE, M. (1957), "In Defense of the Number Two", *Studies Presented to Joshua Whatmough* ('s-Gravenhage, Mouton), pp. 64-72.

HARARY, F. and PAPER, H. H. (1957), "Toward a General Calculus of Phonemic Distribution", *Language,* Vol. 33, pp. 143-69.

HARRIS, Z. S. (1952), "Discourse Analysis", *Language,* Vol. 28, pp. 1-30.

HARRIS, Z. S. (1957), "Co-occurrence and Transformation in Linguistic Structure", *Language,* Vol. 33, pp. 283-340.

HARWOOD, F. W. (1955), "Axiomatic Syntax - The Construction and Evaluation of a Syntactic Calculus", *Language,* Vol. 31, pp. 409-15.

HERDAN, G. (1956), *Language as Choice and Change* (Groningen, Noordhoff).

HERDAN, G. (1960), *Type-Token Mathematics* ('s-Gravenhage, Mouton).

HJELMSLEV, L. and ULDALL, H. J. (1957), *Outline of Glossematics* (Copenhagen, Nordisk Sprog- og Kulturforlag).

HOCKETT, C. F. (1954), "Two Models of Grammatical Analysis", *Word,* Vol. 10, pp. 210-34.

HOIJER, H. (1956), "Lexicostatistics: A Critique", *Language,* Vol. 32, pp. 49-60.

HYMES, D. H. (1959), "Genetic Classification: Retrospect and Prospect", *Anthropological Linguistics,* Vol. 1, No. 2, pp. 50-66.

Information and Control (journal) (1957-) (New York, Academic Press).

JOOS, M. (1950), "Description of Language Design", *Journal of the Acoustical Society of America,* Vol. 22, pp. 701-8.

JOSSELSON, H.H.(1953), *The Russian Word Count* (Detroit, Wayne University Press).

KELLER, K.C. and SAPORTA,S. (1957), "The Frequency of Consonant Clusters in Chontal", *International Journal of American Linguistics,* Vol. 23, pp. 28-35.

KOSCHMIEDER, E., (1956), "Die Mathematisierung der Sprachwissenschaft", *Forschungen und Fortschritte,* Vol. 30, pp. 210-16.

KROEBER, A. L. (1955). "Linquistic Time Depth Results So Far and their Meaning", *International Journal of American Linguistics,* Vol. 21, pp. 91-104.

KROEBER, A. L. and CHRÉTIEN, C. D. (1937), "Quantitative Classification of Indoeuropean Languages", *Language,* Vol. 13, pp. 83-103.

Кулагина, О. С. (1958), "Об способе определения грамматических понятий на базе теории множеств", ПРОБЛЕМЫ КИБЕРНЕТИКИ 1., (Москва, Государственное издательство физико-математической лиературы), pp. 203-14.

LAMBEK, J. (1958), "The Mathematics of Sentence Structure", *American Mathematical Monthly,* Vol. 65, pp. 154-70.

LEES, R. B. (1955), "The Basis of Glottochronology", *Language,* Vol. 29, pp. 113-27.

LEES, R. B. (1957), "Review of Chomsky, *Syntactic Structures", Language,* Vol. 33, pp. 375-408.

LEES, R. B. (1959), "Automata and the Generation of Sentences", *Anthropological Linguistics,* Vol. 1, No. 4, pp. 1-4.

LUHN, H. P. (1958), "The Automatic Creation of Literature Abstracts", *IBM Journal of Research and Development,* Vol. 2, p. 159.

ŁUKASIEWICZ, J. (1957), *Aristotle's Syllogistic* (Oxford).

MANDELBROT, B. (1954), Structure Formelle des Textes et Communication: deux Études", *Word,* Vol. 10, pp. 1-27.

MANDELBROT, B. (1955), "On Recurrent Noise Limiting Coding", *Symposium on Information Networks, Proceedings 1954* (New York, Polytechnic Institute of Brooklyn), pp. 205-21.

MANDELBROT, B. (1957), "Linguistique Statistique Macroscopique", in Apostel, L., Mandelbrot, B., and Morf, A., *Logique, Langage, et Théorie de l'lnformation* (Paris, Presses Universitaires de France).

MANDELBROT, B. (1959), "A Note on a Class of Skew Distribution Functions: Analysis and Critique of a Paper by H. A. Simon", *Information and Control,* Vol. 2, pp. 90-9.

Mathematical Linguistics (Journal) (1957-) (Tokyo, National Language Research Institute).

Mechanical Translation (MT) (1954-), (Journal) (Cambridge, Mass., *M.I.T.).*

MENZERATH,P. (1950), "Typology of Languages", *Journal of the Acoustical Society of America,* Vol. 22, pp. 698-701.

MENZERATH, P. and MEYER-EPPLER, W. (1950), *Sprachtypologische Untersuchungen, I* (Lund).

MEYER-EPPLER, W. (1955), "Anwendung der Kommunikationsforschung auf lautsprachliche und typographische Probleme", *Sprachforum,* Vol. 1, pp. 70-7.

MEYER-EPPLER, W. (1959), *Grundlagen und Anvendungen der Informationstheorie* (Berlin-Göttingen-Heidelberg, Springer).

MILLER, G. A. (1956), "Human Memory and the Storage of Information", *IRE Transactions on Information Theory,* Vol. IT-2, pp. 129-37.

MILLER, G. A. (1957), "Some Effects of Intermittent Silence", *American Journal of Psychology,* Vol. 70, pp. 311-3.

MILLER, G. A. and FRIEDMAN, E. A. (1957) "The Reconstruction of Mutilated English Texts", *Information and Control,* Vol. I, pp. 38-55.

MILLER, G. A. and NEWMAN, E. G. (1958), "Tests of a Statistical Explanation of the Rank-Frequency Relation for Words in Written English", *American Journal of Psychology,* Vol. 71, pp. 209-18.

MILLER, G. A, NEWMAN, E. G. and FRIEDMAN, E. A. (1958), "Length-Frequency Statistics for Written English" *Information and Control,* Vol. 1, pp. 370-89.

OETTINGER, A. G. (1957), "Linguistics and Mathematics", *Studies Presented to Joshua Whatmough* ('s-Gravenhage, Mouton,) pp. 179-86.

OETTINGER, A. G. (1960a), *Automatic Language Translation: Lexical and Technical Problems,* Harvard Monographs in Applied Science, No. 8 (Cambridge, Mass., Harvard University Press).

OETTINGER, A. G. (1960b), "Automatic Syntactic Analysis and the Pushdown Store", *Proceedings of the Symposia in Applied Mathematics,Vol.* 12 (Providence, American Mathematical Society) (in press).

Patent Office (1957-), *Research and Development Report* (U.S. Dept. of Commerce).

Preprints of Papers for the International Conference on Scientific Information, (1958), (Washington, D. C., National Research Council).

REA, J. A. (1958), "Concerning the Validity of Lexicostatistics", *International Journal of American Linguistics,* Vol. 24, pp. 145-50.

RHODES, I. (1959a), "A New Approach to the Mechanical Translation of Russian", (Washington, D. C., National Bureau of Standards, unpublished report).

RHODES, I. (1959b), "A New Approach to the Mechanical Syntactic Analysis of Russian", (Washington, D. C., National Bureau of Standards, unpublished report).

ROSS, A. S. C. (1950), "Philological Probability Problems", *Journal of the Royal Statistical Society,* Vol. 114, Series A, pp. 19-59.

SAPORTA,S. (1957), "Methodological Considerations Regarding a Statistical Approach to Typologies", *International Journal of American Linguistics,* Vol. 23, pp. 109-13 (1957).

SGALL, P. (1959), "Nové otázky matematických metod v jazykovĕde", *Slovo a Slovesnost,* Vol. 20, pp. 44-55.

SHANNON, C. E. (1951), "Prediction and Entropy of Printed English", *Bell System Technical Journal,* Vol. 30, pp. 50-64.

SHANNON, C. E. and WEAVER, W. (1949), *The Mathematical Theory of Communication* (Urbana, University of Illinois Press).

SHERRY, M. E. (1960), "Syntactic Analysis in Automatic Translation", *Mathematical Linguistics and Automatic Translation,* Rpt. No. NSF-5, Harvard Comp. Lab.

SHERRY, M. and OETTINGER, A. G. (1960), "A New Model of Natural Language for Predictive Syntactic Analysis" (to appear in *Proceedings* of the Fourth London Symposium on Information Theory, Sept. 1960).

SIMON, H. A. (1955), "On a Class of Skew Distribution Functions", *Biometrika,* Vol. 42, pp. 425-40.

SOMERS, H. H. (1959), *Analyse Mathematique du Langage* (Louvain).

STUMPERS, F. L. (1953), "A Bibliography of Information Theory-Communication Theory-Cybernetics", *IRE Transactions on Information Theory,* Vol. PGIT-2. First Supplement in Vol. IT-1, No. 2 (Sept., 1955), pp. 31-47. Second Supplement in Vol. IT-3 (June 1957), pp. 150-66.

SWADESH, M. "Salish Internal Relationships", *International Journal of American Linguistics,* Vol. 16, pp. 157-67.

SWADESH, M. (1955), "Towards Greater Accuracy in Lexicostatistic Dating", *International Journal of American Linguistics,* Vol. 21, pp. 121-37.

Symposium on the Structure of Language and its Mathematical Aspects (1960), *Proceedings of the Symposia in Applied Mathematics,* Vol. 12 (Providence, American Mathematical Society) (in press).

UNGEHEUER, G. (19593), "Das logistische Fundament binärer Phonemklassifikationen", *Studia Linguistica,* pp. 69—97.

UNGEHEUER, G. (1959b), *Logischer Positivismus und moderne Linguistik (Glossematik)* (Uppsala, Almqvist and Wiksell).

WELLS, R. (1947), "Immediate Constituents", *Language,* Vol. 23, pp. 81-117.

WHATMOUGH, J. (1956), *Language* (London and New York), chapter XI.

WHATMOUGH, J. (1956), *Poetic, Scientific, and Other Forms of Discourse* (California University Press): passim.

WHATMOUGH, J. (1957), 'Mathematical Linguistics", *Proceedings of the VIII International Congress of Linguists* (Oslo) (1957).

WIENER, N. (1948), *Cybernetics or Control in the Animal and the Machine,* (New York, Wiley).

WORTH, D. S. (1958), "Transform Analysis of Russian Instrumental Constructions', *Word,* Vol. 14, pp. 247-90.

WUNDHEILER, L. and WUNDHEILER, A. (1955), "Some Logical Concepts for Syntax", in Locke, W. N., and Booth, A. D. (eds.), *Machine Translation of Languages* (New York, Wiley).

YNGVE, V. H. (1955), "Sentence-for-sentence Translation," *Mechanical Translation (MT),* Vol. 2, pp. 29-37.

YNGVE, V. H. (1960), "A Model and an Hypothesis for Language Structure", *Proceedings of the American Philosophical Society,* Vol. 104, pp. 444-466.

YULE, G. U. (1944), *The Statistical Study of Literary Vocabulary* (Cambridge, Cambridge University Press).

ZIPF, G. K. (1949), *Human Behavior and the Principle of Least Effort* (Cambridge, Mass., Addison-Wesley).