Proc. 5th int. Congr. phon. Sci., Münster 1964, pp. 43–60
(S. Karger, Basel/New York 1965).

# Code Theory and 'Discrete Mathematics' in Phonology

By Henning Spang-Hanssen, Copenhagen

Language is public property. This is true of language as an instrument of thought and communication and of language as a source of emotional, e.g. aesthetic impressions. However, it seems to hold true of language even as an object of research. Apparently there is no end to the list of disciplines and sciences which lay claim to language as a genuine part of their field of study. Among other things language has been appointed a set of social habits, a kind of individual behavior, a logical – though perverted – structure (a calculus), a hierarchy of mathematical relations, and a – rather irrational – code. Accordingly, linguistics has been considered part of anthropology, psychology, logic, mathematics, and information theory or communication theory.

Whether language *is* a behaviour or a code, etc., and whether linguistics *is* part of logic or mathematics, etc., are to a considerable extent analytical questions, to be answered differently according to adopted definitions of the terms in question, including the word *is*. Thus the various statements concerning the nature of language may all be true at the same time, on condition that they are individually interpreted as "Language *also* is . . .". And in fact it is very unlikely that all empirical linguistic phenomena can adequately be studied within the frames of any one existing discipline or science. With regard to phenomena of language expression this fact has been duly recognized in the naming of the series of congresses, of which this forms the fifth, as Congresses of Phonetic *Sciences*.

It is definitely to be hoped that still new disciplines by laying claim to language will contribute to the study of language. On the other hand, it sometimes happens that a new claim for some time attracts attention and meets with approval to a degree that is hardly

motivated by the actual gain of knowledge likely to be achieved from this new point of view. To keep up with fashion is a factor of some importance even in linguistic research. In particular it may often reasonably be asked whether a new claim or approach is anything but a new terminology. This question, though sceptical, may not be depreciatory, as the development of linguistic terminology forms an important part of the development of linguistics and of the relations between linguistics and other fields of knowledge. But in order to further the study of linguistic phenomena new claims and statements concerning the nature of language – or of linguistics – must at any rate describe what is less known ("language" or the like) by what is better known, or expound what is complex and less intelligible by what is simple and more easy to understand. In this respect the above statements, as usually presented, leave much to be desired.

"*Linguistics is a discrete mathematics*"

Our present concern is with the claim that language is (also, or partly) a code, and the claim that linguistics is (also, or partly) a discrete (discontinuous) mathematics [cf., for instance, *Martin Joos'* paper in JASA *22:* 701–708 (1950)]. These statements are related to some extent, cf. later; at the outset, however, they will be discussed separately.

As regards the relationships between linguistics and discrete mathematics I may refer to the discussion I have given in a paper read to The 9th Internat. Congress of Linguists (Preprints Cambridge, Mass, pp. 133–138, 1962; Proceedings 724–730). My main points are the following:

1. Qualitative linguistics (as distinct from research including statistical or other quantitative aspects of language) may be called mathematical, in the sense of *axiomatic* (i.e. making use of some explicit model containing axioms or postulates). But as various axiomatic models since long have been developed within linguistics proper, independently of mathematics as a particular science, an extension of the term 'mathematical', so as to cover what is usually called structural linguistics, only seems to blur a useful distinction (between mathematical and axiomatic) by putting a "mathematical cross" on every axiomatic model in any field of study. Thus in this respect no clarity is achieved by describing linguistics as a (discrete) mathematics.



2. The statement "Linguistics is a discrete mathematics" may be read in a *normative* sense, claiming that linguistic models ought to be models developed within mathematics. The part of discrete mathematics most often suggested to deal with linguistic matters – at least in the field of phonetics and phonology (or phonemics) – is *set theory* (Mengenlehre, teorija množestv, théorie des ensembles). In fact this discipline is often regarded as part of logic (symbolic logic) or as (part of) mathematical logic or logical algebra.

*Set-theoretical models in phonology*

Detailed phonological (phonemic) studies applying set-theoretical models are those of *Harary* and *Paper* [Language *33:* 143–169 (1957)] and of *Sigurd* and *Gårding* [Studia Linguistica, pp. 8–34 (1955)]. These papers lack nothing in clarity, and they form important contributions to the particular linguistic matters dealt with. But at the same time they clearly illustrate the limitations of set theory. In my above-mentioned paper I have pointed to certain inadequacies of set-theoretical models in relation to linguistic matters:

a) The notion of *order* is a fundamental in set theory and in the adjoining theory of relations. Thus a fundamental distinction is made between symmetric relations (i.e. sets of combinations in which for any pair of elements x and y both the combination xy and the combination yx are found) and asymmetric relations (in which at least one combination xy has no counterpart yx). But since linguistic conditions of combinability are in important cases neutral as regards order (cf. the combinatory difference between vowels and consonants, or relationships such as concord and government), models introducing order as a necessary notion may have the effect of a strait-jacket. No doubt set theory can be modified in this respect, but not without consequences to other parts of the theory (in particular to the relational property called transitivity), and in any case this modification forms an illustration of a necessary adaptation to linguistic conditions.

b) More serious inadequacies of set-theoretical models are due to their *synthetic* nature. In set theory a set is defined by its elements, and a set of combinations by the collection of *given* pairs of elements. Accordingly, a set-theoretical model may well serve the purpose of describing empirical phoneme combinations by their constituent

phonemes, but it is not particularly suited for classificatory analysis of phonemes in terms of combinatory possibilities (combinability). "It is our aim to provide a technique for describing and quantifying phonemic interaction, thereby sharpening the concept of distribution. In none of this is there any claim that analysis will be aided, although, to be sure, we know of no other adequate technique for handling distribution as rigorously or as effective." [*Harary* and *Paper*, Language *33:* 145 (1957).]

Whether distribution may be handled by other techniques, is partly a question of what is understood by *distribution*. By reference to my report on this subject to The 8th Internat. Congress of Linguists (Oslo 1957) it may be said that when distribution is regarded as a basis of classificatory phonemic analysis, other models, in particular those based on the glossematic concept of relation (or function), seem to be more adequate than set-theoretical models. It is interesting – and promising indeed – that the glossematic typology of relations may be mapped on the set-theoretical properties of relations, and vice versa. As an illustration one may imagine an inventory of "syllables" or "word expressions" like that given in the diagram.

| | b | d | f | g | h | a | e | i | o |
|---|---|---|---|---|---|---|---|---|---|
| b | | | | | | ba | be | | |
| d | | | | | | da | | | do |
| f | | | | | | | fe | fi | fo |
| g | | | | | | | ge | | go |
| h | | | | | | | | hi | |
| a | | | | | | | | | |
| e | | | | | | | | | |
| i | | | | | | | | | |
| o | | | | | | | | | |

In set-theoretical terms this inventory forms an intransitive relation: Two elements (e.g., *a* and *e*) that are found in combination with one and the same element (*b*) are never found in mutual combination. In glossematic terms this inventory is analyzable by means of the relation (function) called solidarity, in that two solidary classes (categories) of constituent parts are recognizable: A member of the class *b, d, f, g, h* never occurs without being accompanied by a member of the class *a, e, i, o* – and vice versa.



In this particular example it seems a pure matter of taste which model to prefer, but in cases of partly transitive sets (e.g., if even the combination *ae* has occurred) the latter model is superior as regards classificatory power (cf. my "Probability and Structural Classification", Copenhagen 1959), and moreover it permits handling the question of accidental gaps versus excluded combinations.

c) With regard to this and other questions of chance set-theoretical models are unsuitable – in accordance with their nature of being defined by *given* sets of combinations. This fact forms a serious restriction to the application of such models in phonology and, considering the development of linguistic analysis towards problems of grammaticalness and generative grammar, even in other linguistic areas. Models of quite a different nature are needed to deal with the prognostic, extrapolating, and generalizing aspects of linguistics, including even the characterization of language as such. "Woher aber kommt dieses *Geltenmüssen* für nicht untersuchte Sprachen: noch nicht untersuchte oder prinzipiell nicht untersuchbare (weil sie nämlich noch gar nicht existieren)? Ein solches 'Muß' ist ein Fremdkörper in der von Bopp, Rask, Grimm begründeten Sprachwissenschaft..." (*Eberhard Zwirner*[19], p. 136).

The considerations relevant to such questions are not to be found in mathematics of the type meant when linguistics is spoken about as a discrete mathematics. With regard to these questions interest is focussed on the calculus of probability and mathematical statistics.

Summing up it may be said that with regard to linguistics set theory is definitely more than a new terminology, but on the other hand the descriptive power of such models is limited. Until now their chief merit is to have thrown light on certain pecularities of *linguistic* problems.

This conclusion seems to be in accordance with the aim of other applications of set theory (or symbolic logic) to phonological problems. *J. Cantineau* [Word *11:* 1–9 (1955)] has compared the set-theoretical notions of relation with *Trubetzkoy's* system of phonological oppositions (Lit. 16, pp. 68–99). *G. Ungeheuer* [Studia Linguistica, pp. 69–97 (1959)] applies logical algebra to "1. die formalen Prinzipien einer binären Klassifikation von Schallereignissen" (in terms of distinctive features); "2. die formalen Prinzipien der analytischen Transkription, die auf dieser Klassifikation aufbaut".

In these papers logical algebra is applied to models already set

up within linguistics, in order to formalize and to develop more consistently the bases of classification and reduction (cf. also *W. Plath's* comments – in Lit. 11, p. 39f. – on *Ungeheuer's* paper). No doubt the conceptual basis of phonological description will benefit by such attempts (quite recent contributions are found in "Problemy strukturnoj lingvistiki", Moskva 1963). However, important basic problems are beyond their compass; among other things the description of phonemes by distinctive features gives rise to the problem of where to end the analysis. If the ultimate elements of the expression structure of a (spoken) language are a set of distinctive features, and if structure is defined as (a hierarchy of) rules of combination (or more precisely: of combinability), it will be necessary to account for the empirical fact that the number of phonemes is – usually, and possibly always – smaller than the number of imaginable combinations of features. This can be done *either* by stating rules governing the combinability of features, *or* by describing missing combinations as accidental gaps.

In the latter case the empirical inventory of phonemes cannot be regarded as exhaustive, i.e. as a structural fact about the language in question. If one is unwilling to accept this consequence, i.e. if one insists on regarding phonemes as structural units, the distinctive features cannot be preserved as structural elements of the language in question; they may, however, be regarded as elements belonging to a certain structure of expression *manifestation*, and as such they may be relevant to more than one empirical language.

Which one of the alternative ways of description to be chosen, i.e. whether to carry on the structural analysis of a language expression below the level of phonemes or not, will mainly depend on the actual possibility (for the language in question) of ascertaining rules governing the restricted combinability of distinctive features. This is a problem of the type accidental gap versus excluded occurrence, and – as mentioned earlier – set theory does not provide tools for solving such questions.

The problem just discussed is reflected in the following quotation from *Jakobson* and *Halle*[4] (p. 217): "... this code includes all the distinctive features to be manipulated, all their admissible combinations into bundles of concurrent features termed 'phonemes', and all the rules of concatenating phonemes into sequences...". It will be seen that no mention is here made of *rules* of concatenating or combining the distinctive features into phonemes.



Another point to notice in this quotation is the use of 'code' to mean linguistic inventories and rules. This brings us to the discussion concerning language and code.

*What is a code?*

All of us should like to know what language is, but if we look for an answer in the statement "Language is a code" we shall be disappointed. For opinions differ as much about the meaning of 'code' as about the meaning of 'language'. Since this paper deals with code in relation to language (in particular: to language expression), an attempt at a general survey of code theories shall not be made. With regard to linguistic applications it is hardly possible to distinguish between communication theory and information theory; in fact, only a few experts seem to manage this distinction (cf. the similar situation of logic and mathematics sharing set theory).

Code is a basic concept in information and communication theory. It might be regarded as an indefinable, and in fact no definition of 'code' is found in *W. Meyer-Eppler's* detailed and comprehensive "Grundlagen und Anwendungen der Informationstheorie" (1959). However, in other works various types of definition or characterization of 'code' are met with:

I. "Quand on parle de code, nous pensons souvent à des secrets ou à des intrigues internationales, mais dans ce livre nous emploierons ce mot dans un sens beaucoup plus général. Tout système de symboles qui, par convention préalable, est destiné à représenter et transmettre l'information entre la source et le point de destination sera appelé un code. Ainsi, en ce sens, la langue française est un code et la langue allemande un autre." *G. A. Miller* (p. 14 of the French edition of "Language and Communication"). According to this kind of definition the statement "Language is a code" becomes trivial: it holds true because code has been defined so as to include language. But with regard to linguistic research this conception of code amounts to the non-trivial hypothesis that language and (other) symbol systems having a communicative function may profitably be studied together. Or in a normative version: Language and (other) communicative systems of symbols ought to be studied together.

II. A similar, yet different conception of code is connected with the view that language should be studied together with (other) symbol systems not because of a common communicative function but because of common formal features. In other words: Language and

certain (probably not all) symbol systems are assumed to be of the same internal nature, and 'code' may be used as a designation of any system of that nature. Opinions differ, however, as to the nature of this nature: Which formal features or characteristics are to form the basis of such a notion 'code'?

a) The characteristic of a code may be sought in the existence of rules governing the combination of elements; thus *combinability* (sočetaemost') is the basic feature. Cf. by *E. V. Padučeva*[10] (pp. 114, 115 transl.): "The resemblance of language to a code is above all based on the fact that a description of the combinability of elements plays an important part both in technical codes and language... The description of a language from the point of view of the combinability of its units will be labelled 'description of a language as code'." As she also remarks, this conception of code basically coincides with a description in terms of distribution. Thus, as regards the qualitative aspects of distribution, reference may be made to the earlier discussion on mathematical (set-theoretical) models applied to phonology; the more so, as in information theory, etc. the description of combinatory conditions of codes (linguistic and non-linguistic) usually is given in such mathematical terms.

In information theory, however, the qualitative description is supplemented by quantitative, in particular statistical points of view, and therefore the above-mentioned conception of code includes the hypothesis that the *quantitative* aspects of combinatory conditions in language may profitably be studied together with combinatory conditions found in (other) symbol systems. Cf. later.

b) In discussions concerning the nature of 'le signe linguistique' attention is often focussed on *arbitrariness of signs* as a characteristic of language and of certain other sign systems. *Saussure*, having suggested "une science qui étudie la vie des signes au sein de la vie sociale ... nous la nommerons sémiologie"[14] (p. 33), describes the main subject of this science as "l'ensemble des systèmes fondés sur l'arbitraire du signe"[14] (p. 100).

Besides language (la langue) he mentions among other examples the system of military signals as belonging to the field of semiology. Even though the designation 'code' was not used by *Saussure*, certain aspects of modern code theory (or theories) have a striking resemblance to his idea – which is not tantamount to saying that code theory has been influenced by *Saussure's* suggestion. Among linguists *Saussure's* idea of a semiology has not found wide



acceptance; the designation has survived, in particular in works by *Buyssens* and by *Hjelmslev*, but there it covers other, and divergent, conceptions as to the basis on which language ought to be studied together with other sign or symbol systems (a more detailed discussion is found, e.g., in my "Recent Theories...[15]).

c) Since all the technical codes in which modern code theory originates are systems of arbitrary signs or symbols, arbitrariness does not form a particularly interesting formal feature from the point of view of code theory, and the claim for language does not in particular seem to be motivated in the arbitrariness of language signs. On the contrary, language (together with certain sign systems) may be found terminologically distinguished from code just by reference to difference of *origin*, in some way related to the question of arbitrariness; "... we distinguish sharply between *language*, which is developed organically over long periods of time, and *codes*, which are invented for some specific purpose and follow explicit rules". *Colin Cherry*[1] (p. 7).

However, this basis of distinction does not seem important to the question of whether language may profitably be studied together with other sign systems. For an organically developed sign system, e.g. the decimal cipher system, may well be studied together with "invented" systems, e.g. the binary or the tetradic cipher system.

Speaking of a binary code it shall be mentioned that the present paper does not deal with the much-debated question whether linguistic description – or if one prefers: the linguistic metalanguage or metatext – is profitably constructed in terms of some particular code, e.g. a binary code. The discussion in this paper is confined to the question of to what extent the study of language phenomena can be furthered by comparing language with other systems of symbols or signs.

III. According to still another conception a code may be defined as a communicative system manifested in a particular way (i.e. in a particular expression substance, or by a particular use of expression substance). This definition or characterization covers among other things flag codes and secret codes expressed by ciphers or by letters occurring with unusual values; moreover, certain ways of symbolization introduced for scientific, technical, or commercial purposes.

By this kind of definition nothing is said about the structure of

the system manifested; hence a code (in this sense) may be identical as to structure with some sign system manifested in a usual or "normal" way, for instance a language. And since by definition the manifestations of such codes deviate from those of normal languages, the study of language expression is not likely to be furthered by the study of codes in this sense – exception made for the very fact that one and the same structure may be manifested in different ways.

A variant of this conception of code is to characterize a code as being *secondary* to some other communicative system of signs or symbols. This kind of definition is also discussed by *E. V. Padučeva* (op. cit., p. 115). But apart from facts of historical origin the question of what is secondary and what is primary does not basically apply to systems but to *texts* (messages) or to a particular *process* of communication. For instance, a secret code or a technical code will be secondary to the original message (in plain text) with regard to the process of encoding, but the reverse holds true with regard to the process of decoding (deciphering).

Here, and in general, the linguistic distinction between text (utterance, message) and system (language) proves essential to the application of code theory to linguistic matters. In the literature on information or communication theory the word code is not rarely used indiscriminately in both ways, thereby causing confusion about the nature of coding as a link in a process of communication. A text may be converted into another version, but not into a system; thus a message cannot be converted into a code in the sense of a system.

IV. The confusion becomes even greater since in the relevant literature 'code' is often *defined as a transformation*, 'eine Zuordnung', or the like, i.e. as a system for converting messages, whereas the *examples* of codes given by the same authors point to the conception of a code as a system of symbols or signs. "... a code is an agreed transformation, usually one to one and reversible, by which messages may be converted from one set of the signs to another. Morse code, semaphore, and the deaf-and-dumb code represent typical examples." (*Colin Cherry*, op. cit., p. 7.) "Unter einem Code versteht man eine Zuordnung zwischen zwei Listen von Zeichen oder Zeichenserien; ... Ein ... Code ist das indoarabische Zahlensystem, das allen möglichen Zahlen eine Serie zuordnet, die aus den 10 Ziffern von 0 bis 9 entnommen ist; ... Das normale Alphabet ist ein Code mit einer Liste von (z.B.) $k = 26$ Zeichen." (*Heinz Zemanek*[18], p. 30.)



Here a clear distinction between code, coding (encoding, decoding, recoding), and the result of coding would be useful. At any rate rules for converting from one system to another must not be identified with the systems; the system (the rules) used for converting a decimal number (say 9) into a binary number (viz. 1001) is not identical with the system of decimal numbers, nor with that of binary numbers; rules for transcribing written Russian into a sequence of Latin characters do not form part of the Russian alphabet (or graphemic structure), nor of the Latin alphabet.

Language can hardly be regarded as a code in the sense of a system of rules connecting one system of signs with another system of signs. It is worthy of mention, however, that *L. Zabrocki*[17] (pp. 64 to 73) interprets the structural hierarchy of any language as a code, i.e. as a coding process proceeding from 'Lautgefüge' to 'Wort', from 'Wortgefüge' to 'Satz', etc. He expressly points to the duality of his concept 'code': "Der *Sprachkode* ist im Grunde genommen ein Transponierungsprozeß. Er enthält zugleich die Gesetze der Transponierung" (p. 73). It is difficult to see, however, whether *Zabrocki's* theory of language as a code is compatible with the conceptions of code developed by information theory etc.

Processes in which two linguistic systems of signs take part are sometimes talked about as being of the same nature as the transformations effected in technical coding processes. "According to our definition, transforming a printed message into Morse code, transliterating from the Cyrillic to the Roman alphabet, enciphering for cryptographic purposes, and replacing decimal numerals by binary numerals belong to one family with translating *Macbeth* into German ..." *A. G. Oettinger*[9] (p. 104). Since, however, the crucial point of machine translation is whether the relation between texts in different languages can actually be substituted by a set of correspondences of the one-to-one type characteristic of technical coding, it seems premature to regard translation on a par with obvious coding processes like transliteration. Cf. also *A. A. Reformatskij's* criticism[12] (pp. 208–215) of the term 'recoding' (perekodirovanie) used of the relation between normal and tactile language as well as of e.g., transliteration.

*Phonological applications of code theories*

Summing up this discussion on the ways in which code is conceived with regard to language, it may be said that for the time

being the notion of 'code' is too complex or too vague to serve as a basis of a uniform study of linguistic matters. Accordingly, papers presented as applications of code theory to language may have little in common – except for the very term "code".

In certain papers code theory seems to be little more than a new terminology applied to considerations built on combinatorics – a discipline known through centuries, and in a variety of sciences (classical probability, genetics, etc., even linguistics) – and to reflections on efficiency, e.g. in linguistic change, previously discussed by linguists (among others by *Otto Jespersen*). It should be noticed, moreover, that information theory etc. usually narrows down the problems of linguistic efficiency to the question of how texts (messages) are efficiently *communicated.* This narrowing down is quite natural – and in fact necessary – from the point of view of the technical applications (telecommunication) in which information theory originates.

Due to its technical perspectives information theory is *normative* and *evaluating:* Code systems are studied with the aim of achieving the greatest possible efficiency, such as saving time or equipment, and as avoiding disturbances. This kind of approach is extremely important, but it does not exhaust the study and the description of language expression, not even with regard to questions of efficiency. Conditions of *acquiring* a code or a language (in childhood or by conscious learning) form a different basis of evaluating expression systems; and basically different from the question of transmitting given information is also the question of permitting new information to be formed and expressed.

The latter question is, among other things, bound up with the possibilities of introducing new word expressions, e.g. admissible but hitherto not exploited syllables, on the basis of a given inventory of phonemes. The possibility of creating new words, and in general of creating new combinations of elements and units, seems to be a characteristic of natural languages. From this point of view the conditions of language expression prove to be more complex than can adequately be dealt with in notions such as 'rationelle Sprache' in the sense suggested by *W. Fucks.* "Wir betrachten eine Sprache (im allgemeinsten Sinn des Wortes), in der zusammenhängende Symbolaggregate (Komplexionen) von maximal *n* Elementen vorkommen. Diese Komplexionen sollen *Wörter* genannt werden, ohne daß diese Bezeichnung mehr als eine formale Ähnlichkeit mit den



Wörtern einer Nationalsprache ausdrücken soll. Ein erstes Charakteristikum einer solchen Sprache ist ihr *Wortvorrat* oder *Vokabular,* d.h. die Zahl der aus $\zeta$ verschiedenen Symbolen durch Bildung von bis zu *n*-stelligen Komplexionen zu gewinnenden verschiedenen Wörtern. Eine Sprache, bei der diese Wortbildungsmöglichkeit restlos ausgenutzt wird, heiße nach *Fucks* eine *rationelle Sprache.*" *W. Meyer-Eppler*[7] (p. 86).

Due regard should be payed to the reservations found in this quotation, but on the other hand it is hard to see how code theory may at all contribute to the study of language structure, if a formal similarity (eine formale Ähnlichkeit) is not considered a sufficient basis. At any rate the similarity has been deemed close enough to justify an adoption of the designations Wort and Sprache.

But in the first place it applies that what according to the above definition is rational, may in another respect appear irrational; this fact is duly recognized elsewhere in information theory, stressing the importance of redundancy as a safeguard against mistakes. Secondly, it should be noted that the failure of a natural language to fulfil the conditions for being 'eine rationelle Sprache' may be due to factors of two different kinds: It may be due to properties of the structure (in that certain combinations – 'Komplexionen' – are structurally excluded), or it may be due to conditions of *usage* only (in that not all admissible combinations have empirically occurred as word expressions).

It is true that in a number of applications of code theory this kind of difference is taken into account; various papers of this nature form important contributions to the qualitative and quantitative description of phonemic and graphemic conditions in various languages. But it may well be asked, whether the linguistic perspectives of these contributions exceed the implications of earlier approaches to phonology (or phonemics) and to phonological statistics; in the latter field one may in particular think of papers by *V. Mathesius*[6] and by other Czech phonologists, cf. the survey by *J. Krámský*[5] in Phonetica.

In various respects the descriptive power of code theory with regard to phonology is obviously hampered by pecularities of technical codes. In discussing the question of functional load (exploitation, Belastung) in phonemics, *J. Rischel* [in Statistical Methods in Linguistics *1:* 13–23 (1962)] points to certain difficulties of a description in purely sequential terms, among other things to "the well-known

fact that certain contrasts between phonemes are 'neutralized' under specific conditions, cp. the neutralization of the opposition aspirated stop:unaspirated stop in final position in Danish" (p. 15). "In all contributions to linguistic methodology which build on the basic notions of information theory, it seems to be implied that linguistic sequences are built up of members from one inventory" (p. 16).

*Rischel* proposes "to introduce the concept 'neutralization' in the purely sequential aspect of language as a Markoff-process. This would involve that surely not only the probabilities but the code *inventory* itself varies throughout sequences: *after certain sequences the difference between two elements is neutralized*" (pp. 15–16). It would take us too far to discuss this contribution to code theory, but anyhow *Rischel's* remarks are noticeable by their suggesting, in fact, that in this field codes may profitably be studied together with language, and not the other way round.

An important fact about most – or all – technical codes is their synthetic nature: The code system consists of a *given* number of elements, and of *given* (explicit) rules for combining them into (potential) messages. Owing to this fact discrete signalling systems and their functioning have been studied by information theory without it being necessary to tackle problems of how to find out the (or a) system that corresponds to – "underlies" – a given message. When dealing with the phonemic or graphemic aspect of linguistic messages, code theory in practice draws upon the results of a phonological (phonemic) or graphemic analysis, carried out on a linguistic basis (cf. the situation previously discussed for set-theoretical descriptions of language expression). "Auf welche Weise man die Nachrichtenobjekte im Informationsvolumen" (e.g., the phonemes of an utterance) "erkennen kann, bleibe zunächst offen. Die hierzu geeigneten *Analysiermethoden* bilden eines der schwierigsten Probleme bei der praktischen Anwendung informationstheoretischer Methoden. Wir setzen voraus, die Analyse sei bereits durchgeführt..." *W. Meyer-Eppler*, op. cit., p. 58.

Thus it is an open question whether code theory can furnish a new and more general basis of setting up elements like phonemes. So far there seems to be no theory overbridging the division which *Saussure* introduced at once in the semiology he had suggested himself: "La langue présente donc ce caractère étrange et frappant de ne pas offrir d'entités perceptibles de prime abord, sans qu'on puisse douter cependant qu'elles existent et que c'est leur jeu qui la constitue. C'est là sans doute un trait qui la distingue de toutes les autres institutions sémiologiques" (Cours p. 149).



In the case of continuous signals, including speech in its *physical* aspect, the situation is different. Since the notion of code – whatever particular definition adopted – is always bound up with discrete (discontinuous) signals, code theory becomes relevant to the speech continuum only where attempts are made at quantizing speech into recurrent elements. The background of telecommunication research in this field lies in the importance of compressing speech into signals occupying less channel capacity; in addition to ways of compression by which the speech signal remains continuous, various ways of discrete (parametric) compression, making use of vocoders of different constructions, are being developed (cf. for instance the recent survey by *M. A. Sapožkov*[13]).

It is beyond the scope of the present paper to discuss whether human perception of speech involves neurological processes similar to such technical coding processes – a view advocated in cybernetics. But it is highly relevant to phonology whether it is possible by technical means to quantize speech into code elements corresponding to phonemes. In her report to The 8th Internat. Congress of Linguists (Oslo 1957), *Eli Fischer-Jørgensen*[2] discusses the possibility of obtaining 'phonemes from curves' (by means of "phoneme detectors") but answers in the negative. It is, indeed, unlikely that a machine of human ability in recognizing phonemes will ever be constructed. But it is a matter of human ability in constructing machines, *in what degree* a mechanical segmentation of speech into linguistically relevant elements is possible; among other things, the elements obtained need not be of the same extension as phonemes.

Considering the economical interests attached to this possibility – closely related to conversion of speech into writing – a good deal of effort will probably be devoted to such tasks, and phonology and phonetics may from this research learn something about the nature of phonological analysis, in the same way as research on machine translation has thrown light on certain hitherto unnoticed presuppositions of man-oriented grammar. Anyhow, these perspectives are only by-products of code theory, and they are connected with the conception of language as a code in an indirect way only.

The conclusions arrived at do not raise the expectations from applying code theory to questions of phonology (in the wider sense).

In other linguistic fields the prospects of code theory may be others; but conclusive judgments in the negative occur in the literature: "Two opinions current in MT writings on language are that language is a code and that the code is fundamentally binary. Both these views are, from the standpoint of a communication engineer, tenable and useful. From the linguistic standpoint, however, these views are both questionable and unhelpful; and they have hampered MT work because they misrepresent the functioning of language both in its internal relations and in its relations to non-language" *M. A. K. Halliday*[3] (p. 146).

Anyhow, the question of whether language is a code has important theoretical implications, and even answers in the negative may be useful; for – on condition that a clear definition of code is given – we shall learn something about the nature of language by finding out to what extent language is *not* a code.

### *References*


1. *Cherry, C.:* On human communication (Wiley, and Chapman & Hall, New York/London 1957).
2. *Fischer-Jørgensen, E.:* What can the new techniques of acoustic phonetics contribute to linguistics? Proc. the VIIIth int. Congr. Ling. Oslo 1957, pp. 433–480.
3. *Halliday, M. A. K.:* Linguistics and machine translation. Z. Phon. Sprachw. Kom. *15:* 145–158 (1962).
4. *Jakobson, R.* and *Halle, M.:* Phonology in relation to phonetics, in *L. Kaiser's* Manual of phonetics (Amsterdam 1957).
5. *Krámský, J.:* Quantitative phonemics in the last decade. Phonetica *8:* 166–185 (1962).
6. *Mathesius, V.:* La structure phonologique du lexique du tchèque moderne. TCLP *1:* 67–84 (1929). – Zum Problem der Belastungs- und Kombinationsfähigkeit der Phoneme. TCLP *4:* 148–152 (1931).
7. *Meyer-Eppler, W.:* Grundlagen und Anwendung der Informationstheorie (Springer, Berlin 1959).
8. *Miller, G. A.:* Language and communication (McGraw-Hill, New York 1951; French edition 1956).
9. *Oettinger, A. G.:* Automatic language translation (Cambridge, Mass. 1960).
10. *Akhmanova, O.S.* (red.): O točnykh metodakh issledovanija jazyka (Moskva 1961).
11. Trends in European and American linguistics (Utrecht 1961).
12. *Mološnaja, T. N.* (red.): Issledovanija po strukturnoj tipologii (Moskva 1963).
13. *Sapozkov, M. A.:* Rečevoj signal v kibernetike i svjazi (Moskva 1963).
14. *Saussure, F. M.:* Cours de linguistique générale, 3e éd. (Lausanne/Paris 1931).
15. *Spang-Hanssen, H.:* Recent theories on the nature of the language sign. TCLC *9* (1954).
16. *Trubetzkoy, N. S.:* Grundzüge der Phonologie. TCLP *7:* 68–99 (1939).
17. *Zabrocki, L.:* Sprachkode. Z. Phon., Sprachwiss. Kom. *14:* 64–73 (1961).
18. *Zemanek, H.:* Elementare Informationstheorie (Oldenbourg, Wien/München 1959).
19. *Zwirner, E.:* System der Sprache und System der Wissenschaften. Indog. Forsch. *68:* 133–148 (1963).


Author's address: Dr. H. Spang-Hanssen, Moltkesvej 40, *Copenhagen* (Denmark).



**Discussion**

*Herdan* (Bristol): I do not know who ever described the application of mathematics to language in the words used by the lecturer, namely "Language is a code" and "Language is discrete mathematics". Rightly understood, no such identification is intended by the introduction of mathematical methods in linguistics. Such application has a sound empirical basis. It was simply found that language has certain aspects in common with artificial codes, and since we know the structure of such codes very well, they being of our own making, it is quite natural that we should make use of this when it comes to describing language. This is all it comes to. There is no identification implied in this between natural languages and artificial codes.

Mr. *Spang-Hanssen* said that he did not see the connection between *de Saussure's* teaching and code theory, or that there was no historical connection between the two. He evidently does not take into due consideration de *Saussure's* axiom of the independence of sound and meaning. If this is understood in all its implications it leads to the conception of language being in certain aspects, and specially so on the phonemic and alphabetic levels, very similar to artifical coding systems. In particular, Information Theory which works with the conception of language as a code is only possible if the frequency distribution of phonemes and of alphabetic symbols remains sensibly stable regardless of the content of sufficiently long messages or texts.

As to the statement that "Language is discrete mathematics", if it was ever made like this, it ought not to have been. Although we make extensive use of continuous mathematics in engineering, nobody has ever said that engineering is continuous mathematics. To say so would only provoke the obvious objection that engineering is engineering and mathematics is mathematics. Similarly, it would not make sense to say that language is a discrete mathematics. Both engineering and language are fields in which mathematics can be applied profitably. They are thus fields of applied mathematics, not the mathematics themselves.

*Bés* (Buenos Aires): Je voudrais souligner l'observation faite par M. *Spang-Hanssen*, laquelle s'appuie, à son tour, sur les remarques de M. *Rischel*. Il est bien connu que pour certaines formulations phonologiques on emploie non seulement les unités qui précèdent à un élément donné mais aussi celles qui le suivent dans la chaîne. En espagnol et dans beaucoup d'autres langues, entre voyelle et /p/, la seule nasale qu'on trouve c'est [m]; donc [m] a une information plus faible que si elle était en opposition avec les autres nasales, c'est à dire, quand elle est suivie par une voyelle. Si on exprimait ce fait par le seul recours aux éléments précédents, la formulation qui en résulterait ne serait pas acceptable du point de vue phonologique à cause de sa complexité. On peut donc tirer la conséquence que non seulement il y a des différences d'inventaire dans les différents points de la chaîne mais que celles ci, à son tour, sont parfois déterminées, au moins partiellement, par les phonèmes qui suivent. Cela pose de sérieux problèmes à l'application du procès de Markoff.

*Fry* (London): I would put in a plea for using the expression "artificial code" as Mr. *Herdan* has done in contrast with "natural languages".

This seems to me much clearer than the distinction between codes in which the rules are "explicit" or "implicit", as Mr. *Spang-Hanssen* has characterized them. In an artificial code the rules are formulated before the code is used, and thus we know what the rules are from the start; in natural languages we do not start by knowing the rules and we have to try to discover them. But unless a code is in use, whether it be an artificial code or a natural language, the rules are always implicit.

*Tillmann* (Bonn): Die Frage, ob Sprache ein Code sei oder nicht, kann offenbleiben, wenn man darauf hinweist, daß Sprache immer in einer mittelbaren, sagen wir: codier-

ten Form zugänglich ist. Selbst wenn man es aus wissenschafts- und erkenntnistheoretischen Überlegungen unterlassen muß, der Beziehung zwischen Sprache-an-sich und deren Transformation, d.h. Codierung, und ihrer Beschreibung weiter nachzugehen, so kann meines Erachtens doch die Auffassung der beschriebenen Sprache als Code manche Scheinprobleme auflösen helfen, z.B. das der mono- oder biphonematischen Wertung von Diphthongen und Affrikaten. Ja, man könnte sogar von der «codetheory» Argumente für die Möglichkeit der an sich relativ arbiträren Entscheidung für das eine oder das andere erwarten.