[From: *Information retrieval and machine translation*, ed. Allen Kent. International conference… September 6-12, 1959, Cleveland, Ohio; part 2 (New York: Interscience Publishers, 1961)]

CHAPTER 50

Linguistic Problems in the Design of a Machine Language for the Information Machine*

V. V. IVANOV

Institute of Exact Mechanics and Computing Machinery,
Academy of Sciences, Moscow, U.S.S.R.

1. In the information machine to be designed, the entire body of knowledge related to a certain field of science should be recorded in an abstract machine language. The linguistic problems of the composition of this new language are inseparable from the problem of translation from a given language to an abstract machine language. It has been frequently suggested in the current literature that this process of translation must be accomplished by humans and not by machines. In other words, it has been suggested that it is necessary to record in machine language the entire body of information before it is introduced into an information machine. More recently, however, even those scientists who had earlier supported this point of view have been inclined to accept the premise that this process may be mechanized.† If this phase of the work is to be done by human beings, the use of the information machine would be quite complicated and more expensive than otherwise. It should be recalled that analogous problems arose also during the first stage of machine translation work, when several specialists considered that a text could not

* This article presents the essence of the work performed for the Laboratory of Electro-modeling of the Academy of Sciences of the U. S. S. R. It was presented at the May 28, 1957 session of the conference on the problems of the design and construction of a new information machine with a long-lasting memory. The author wishes to express his gratitude to E. V. Paducheva, M. K. Polivanov and V. A. Uspensky for their valuable remarks during the discussion of the original plan of the project. The Russian text was published in "Papers on machine translation," Leningrad, 1958, p.p. 10-39, the Chinese text in "Yuyanxue Yicong," 1959, 2, pp. 26-32.

† See, for example, J. W. Perry, Allen Kent, and Madeline M. Berry, "Machine Literature Searching," Interscience, New York-London, 1956, pp. 70 and 130.

be introduced into a translating machine without "pre-editing." The situation in the case of the information machine is similar. The difficulty of translating from a natural language to an abstract language by means of human effort does not mean that in the future a specially developed abstract language cannot be used by scientists. It is, therefore, completely within the realm of possibility that not only can a machine (I mean, a self-educating machine) learn human language but also that human beings can learn machine language, if summaries of auto-abstracts were to be written immediately in abstract language by authors in order to simplify the work of the machine. However, this does not mean that the extremely complex problem of recording of all existing information in the scientific literature and of translating it from natural languages to an abstract one can be assigned to human beings. It would require a very great amount of work and a long period of time for human beings, while a machine might accomplish this task in a very short period as compared to that necessary when utilizing human effort. For this reason it is necessary to develop the algorithms of translation from natural languages to an abstract language. One should not postpone the development of these algorithms until the abstract language has been perfected. On the contrary, the linguistic portion of the problem must be considered simultaneously with the mathematical-logical portion of the problem during the development of the abstract language. In order to make possible the translation from a natural language into an abstract language (which will serve for the recording of information in a machine), the later must satisfy certain linguistic requirements. A condition necessary for the application of an abstract language is the possibility of translating texts from natural to abstract languages (and vice versa) (independent of the circumstances of whether this translation will be accomplished by man or by machine). Therefore a linguistic as well as a mathematical-logical way of its construction is possible.

2. When the linguistic requirements for the abstract language are formulated, this language may be considered as a meta-language for natural languages. A language used for the scientific description of a natural language may be called a meta-language; consequently the grammatical rules and vocabulary of a natural language may be considered meta-language used for the description of the language (i.e. set of sentences) of a given text. Therefore the abstract machine language is a meta-language of the second order (meta-meta-language), that is, a meta-language for description of a natural language (or of the set of grammatical rules and vocabularies) which in its turn serves as a meta-language for the formal analysis of scientific and technical texts.* The linguistic (structural) analysis of natural

* In connection with this, it is necessary to stress the importance of the theory of meta-semiotics in Hjelmslev's conception (see B. Siertsema, "A Study of Glossematics," 'S-Gravenhague, 1954, pp. 214-216).

languages may serve as a method of the design of a machine language. The methods which are used in such analyses are analogous to logico-mathematical methods of constructing abstract machine languages as meta-languages (or meta-theories) for a concrete science.

According to Kleene, "from the standpoint of view of meta-theory, the object theory is not properly a theory at all as we properly understood the term, but a system of meaningless objects, like the positions in a game of chess: subject to mechanical manipulations like the moves in chess. The object theory is described and studied as a system of symbols and objects built up out of symbols. The symbols are regarded simply as various kinds of recognizable objects"*: This definition of the task of meta-mathematics corresponds to the principles of modern structural linguistics, which were formulated in de Saussure's theory†.

While applying logico-mathematical and linguistic methods during the construction of the machine language we suppose to get similar results by both procedures because of the similarities of applied methods and because the same object (the scientific language) is investigated. The possibility of construction of an abstract machine language for coding scientific and technical information is based on the fact, which is well expressed by Sapir's words: "The proper medium of scientific expression is a generalized language that may be defined as a symbolic algebra of which all known languages are translations."** The two main methods of design of abstract language (linguistic and logico-mathematical) have the purpose of investigating this "symbolic algebra."

As long as the abstract language is designed on the basis of the formal analysis of natural languages, the meaning of units of abstract language is determined only on the basis of their relations with the units of natural language (and vice versa). The established rules of translation from a natural language to the abstract language (and vice versa) offer a possibility for expressing in the machine language that information which is transmitted by means of the grammatical and lexical code of the object language (that is, the natural language), and which is not redundant from the point of view of the problem to be solved by the information machine. Every natural language is able to transmit various types of information about the various phenomena of reality and their comprehension by human beings. Therefore, the designing of an abstract language which should be adequate for the recording of all information transmitted by natural languages is extremely difficult. In the near future, it would be a

---

*S. C. Kleene, "Introduction to Meta Mathematics," Amsterdam-Groningen, 1952, p. 62.

† See V. V. Ivanov, "Linguistics and Mathematics," in "Bulletin of the Society for the Problems of Machine Translation" (abbreviated "Bulletin"), no. 5, Moscow, 1957, p. 7.

**E. Sapir, "Language," New York, 1943, pp. 223-224.

realistic program to construct an abstract language with information capacity adequate only to that of a natural language of scientific and technical literature. A peculiarity of the latter is that it has an "exactly specified structure,"* that is, for this language the definition of a "marked" class of words and expressions which have sense can be given. This means that a semantic meta-language built for this language can be formalized.

During the automatic analysis of scientific and technical texts put into the machine, the elements to be mapped into the abstract machine language should be found; the elements which belong to ordinary speech and are redundant from the point of view of the goals of the information machine should be separated from the elements of the former type. These elements should not be mapped into the machine language. For this reason, the linguistic and logico-mathematical methods of design of a machine language should give close (if not equivalent) results in the recent stage of development. In case of a mismatch of the results of the linguistic and the logico-mathematical methods of language design, the final form of the meta-language should be got by the intersection of the two models obtained by these methods, in so far as the complement of the logico-mathematical model to a linguistic model may be considered as a semantically redundant part of the former, and the complement of the linguistic model to the logico-mathematical model appears to be untranslatable into natural language. As a next step, the investigation of the mismatching parts of logico-mathematical and linguistic models of the abstract language may be suggested. This would be important, on the one hand, for the experimental design of an abstract language for a broader field of information, and on the other hand, for the modeling of the relations between language and thought. The sequence of the problems to be solved should be preserved. One should proceed from the performance of relatively simple experiments to more complex ones. For the first approximation to an abstract language design, the design of abstract languages for individual fields of science and the simplification of natural languages may be suggested. The languages for individual fields of science and the meta-languages for the simplified (elementary) languages may serve as models for the future work on further development of abstract languages in general.*

3. In order to translate from a natural language to an abstract language (and vice versa) a one-to-one relation should be established

---

*A. Tarski, "The Semantic Conception of Truth and the Foundations of Semantics," in the collection "Semantics and the Philosophy of Language," Urbana, 1952, p. 18-19.

*The elementary linguistic and mathematico-logical problems may be solved by this kind of modeling. In this field, Hjelmslev's works are worth consideration, in which certain principal problems of language structure analysis are solved on the basis of the comparison of languages with the simplest types of semiotic systems.

between the vocabulary of the natural language and the vocabulary inventory schedule (or with the catalog) of concepts of the machine language. This may be accomplished by splitting the meanings of the words of a natural language into elementary meanings; each of them should be mapped into one element of the vocabulary of the abstract language. Such elementary meanings may be called semantemes or semantical distinctive features (elements). The meaning of every lexical unit may be considered as a set of such elements. The set of semantic elements which forms the meaning of the given lexical unit is defined by the number of semantic fields in which this unit is included. The semantic field consists of a set of semantic elements, which are opposed to each other in the system of language (that is, which are paradigmatically different from each other, like the distinctive features of phonemes*). As long as the problem of multiple meaning is mapped into the intersection of these two fields. Formal an example, one of the words of natural language may be selected which is not a term: in modern English the meaning of the word "notorious" is defined by the structure of the semantic field of words to which it belongs; this field indicates reputation, glory, where this word is opposed to the word "famous." At the same time the same word belongs to the semantic field of words related to (negative) qualification of persons, where it is opposed to the word "bad." Consequently the meaning of this word is defined as the set including these two semantic elements. Therefore, in the meta-language its meaning is mapped into the intersection of these two fields. Formal semantic analysis of natural language, following Trier's theory of the linguistic field (anticipated already by Leibniz†, who in this problem, like in many others, was the forerunner of ideas of our modern science) should lead to the design of a meta-language for a given natural language. In accordance with this premise, the semantic meta-language, which is built up on the basis of a formal semantic analysis of natural languages, may be used as an abstract language for machines. The semantic elements in the vocabulary of a machine language should be grouped according to semantic fields (which have been already anticipated in practice in the form of " idioglossaries" for machine translation.)

In addition to the grouping of semantic elements according to semantic fields in Trier's concept, the classification on syntagmatic criterion is also expedient (i.e. a grouping according to Portzig's

* See V. V. Ivanov, "The Concepts of Neutralization in Morphology and Lexics," in "Bulletin," no. 5, Moscow, M. 1957, p. 55-57 (cf. also French translation of this article in "Travaux de 1'Institut de linguistique," vol. 2, Paris, 1957).

† See S. Ohmann, "Theories of the Linguistic Field," in "Word," v. 9, 1953, no 2, p. 130. It should be recalled that Leibniz (like Newton in his early writings studied only recently in 1957) was occupied with the problem of a universal abstract language for science, close to the recent research activity in this field.

concept of a semantic field). This means that in the vocabulary of the abstract language, the peculiarities of syntagmatic relationships in linear sequence of semantic elements should be shown (for example, the relationship between a semantic element expressing the name of a chemical substance, with a semantic element expressing a chemical reaction). For correct analysis of text, it is necessary to fix syntagmatic relationships in the vocabulary. If a multiple-meaning word appears in one of its meanings, then this is usually conditioned syntagmatically; therefore, if one finds the semantic element of one word, the information about the meaning of the related word is readily obtained if the syntagmatic (i.e. linear, function of context) relation between their meanings is obvious. If we continue to consider examples from everyday language, we may use the case which is considered by U. R. Kurylowicz.* The set of semantic elements, which compose the meaning of the word osel (ass) in the Russian language, is defined by the circumstance that this word fits: in the semantic field of animal names (osel ass, verbl'ud camel, etc.); and in the semantic field of words which qualify the mental abilities of human beings (asel ass, durak crazy, etc.). The meaning of the multiple-meaning word can be established easily by the given context (difficulties arise only in connection with the so-called untranslatable "puns," however, in this case, like in other cases of use of words in artistic literature, the violation of the probability laws is obvious, which is caused by the specific role of the act of communication).*

The above discussion was concerned with the multiple-meaning words of the ordinary language. A certain number of such words is frequently encountered in any specialized text (see at least the verbs like to get in English and att fa in Swedish). Even this circumstance alone proves the necessity of the described method of the semantic analysis. However, semantic analysis may also be very useful during the coding of information expressed in single-meaning terms. It proves to be very economical to represent every term of scientific-technical language in the form of the product of several semantic factors (semantic elements). This principle has been proposed by Perry, Kent, and Berry†; it may be clarified in the following example: the term "thermometer" is coded in the abstract machine language as an intersection of concepts (product of semantic factors) "measurement," "instrument," and "temperature"; moreover, an

* U. R. Kurylowicz, "Notes about the Meaning of Words," Problems of Linguistics, 1955, no. 3, p. 18 (to be translated into English in "Word").

* See, the wide application of these methods in the classical Chinese poetry, in the ancient Vietnamese language, etc. These cases (like other generally extraordinary use of words in artistic literature) fit with a statistical definition of information, cf. V. V. Ivanov, linguistic problems of poetical translation, "Abstracts of the conference on machine translation," Moscow, 1958, p. 40.

† See Perry, Kent and Berry, "Machine Literature Searching," Interscience, New York, 1956, p. 78.

"analytical relation" is introduced, i.e. an indication that the given term is connected with "measurement" as a process, with "instrument" as an object included in the given class, and with "temperature" as an object of an action.** Moreover, a discriminating symbol is introduced to differentiate the set of semantic factors from an analogous set for the word "pyrometer." The system suggested by Perry, Kent and Berry hardly can be recognized as fully satisfactory, partly because it does not reflect the fact that the same term has to be resolved into semantic elements individually depending upon the context (i.e., depending upon the relations of a syntagmatic character).* At the same time, the principle for designing a system of semantic products is doubtful, because in this case it is connected neither with the analysis of natural languages, nor with the investigation of "recognizable objects" in the meta-theory of an object science. However, a fruitful idea is doubtlessly presented for the division of the plane of content of scientific language. This idea corresponds to one of the most important achievements of modern structural linguistics, that is the discovery of the fact that every level in a language may be represented as a system of elementary units (phonemes for the phonological level, semantemes for the semantic level, etc.).† The coding of information by means of semantic factors in the information machine facilitates searching and permits the perfection of a search on the basis of new combinations of semantic factors.

Not all of the elements of the semantic vocabulary system of a natural language contain information from the point of view of those problems which are to be solved by the information machine. In particular, stylistic and emotive expressions are superfluous. They may appear even in the words of scientific-technical language (see the emotive value of the words: "cybernetics," "gene," and "chromosome" in certain articles). For the recording of scientific-technical information, the redundant emotive elements of meaning should be completely eliminated in the information machine. The investigations in the field of general semantics may be of *great* interest in this regard despite the fact that the name of this school itself has got in certain articles an evidently emotive value.

** See Perry, Kent and Berry, op. cit., p. 86.

* According to the author, this is the most important argument, conducted (even if it is not formulated linguistically), in the article of Y. Bar-Hillel, "A Logician's Reaction to Recent Theorising on Information Search Systems," in "American Documentation," v. XIII, 1957, no. 2, p. 108. I do not consider it necessary to analyze here the other conclusions made in the article of Bar-Hillel, as long as these conclusions are pertinent only to the automatic search of literature and this paper concerns the automatic analysis and search of scientific-technical information, included in the literature.

† On the mathematical concepts connected with these linguistic ideas, see V. V. Ivanov, "n-Dimension Space of Language" in "Bulletin," 1957, no. 5, p. 51-52.

4. The problem of vocabulary of abstract language has been developed to a more advanced level than the problem of the grammar of abstract language. That which is suggested in the works of Perry, Berry, and Kent in the plane of designation of analytic relations (already mentioned above) and of synthetic relations (which refer to the relation to experience) belong also de facto to lexics, i.e. to the analysis of individual terms of concrete language and their translation to abstract language. In linguistic terminology this corresponds to word derivation and word composition and does not correspond to morphology and syntax. It is to be noted that the problem concerning the interrelation of the grammar of natural language and the grammar and vocabulary of abstract language is very complicated.* When the interrelation of the vocabulary of natural language with that of abstract language is spoken of, it is obvious that units of the vocabulary of natural language are mapped into units of the vocabulary of abstract language. The introduction of "relations" (analytical and synthetic relations) into the former in the works by Perry, Kent, and Berry does not change the essence of the situation because these relations together with the semantic factors form complicated lexical units. (Moreover, it is possible, in contradiction to the system of Perry, Kent and Berry to work with semantic factors alone, without the introduction of relations for the translation of lexical terms). However, if the vocabulary units of natural language are mapped into the vocabulary units of abstract language then this does not mean that, the grammatical units of natural language should be necessarily mapped into the grammatical units of abstract language. On the contrary, it is desirable to have such a system of coding information in the abstract language which permits that a sufficient part of the grammar of the natural language should be translated into the vocabulary of the information machine language, i.e., it should be mapped into the inventory of elementary meanings of the abstract language.*

This is dictated, first by the consideration of engineering aspect of the problem. One of the basic requirements put forward by the engineers is to reduce the number of operations—by the use of a 'memory' having a great capacity. The lexical units (including also complex ones, i.e., the products of semantic factors, if it is expedient to have them in memory as ready made elements) can be fed into the memory once and for all whenever a grammatical rule indicates the

---

* The grammar of abstract language includes logical syntax. The typology of this grammar is related to the problems of rational coding, here not especially considered; however, it can be expected that this language must be analytic.

* R. L. Dobrushin, while discussing this work expressed the thought that machine language may not have, in general, any grammar. This idea is expected to be correct only as far as morphology is concerned. See now on this subject I. A. Mel'chuk "On the "grammatical" part of the interlingua," " Abstracts of the conference on mathematical linguistics," Leningrad, 2959, pp. 60-63.

need for an operation (resulting in combination of elements). Secondly, the necessity of translation of several grammatical rules with vocabulary units is explained by the peculiarities of the grammars of natural languages. Frequently quite concrete meanings are expressed through grammatical methods: it refers to derivational meanings (according to Sapir's classification), for example, the meaning of dimunitiveness, agentiveness, etc. and concrete relational (gender, aspect, number) concepts. It is perfectly clear, that the translation of such grammatical concepts into the grammar of abstract language would make impossible to build a logical syntax for the abstract language. If the vocabulary of abstract language were to be made by linguists, logicians and mathematicians together, as it should be, then during the construction of logical syntax (grammar) for the abstract language, the basic work is assigned to mathematicians and logicians: the linguists' duty is to participate in the criticizing and formalization of the grammar of concrete languages. The grammar of natural languages does not correspond well to the logical syntax of language of science. This concerns not only languages which have grammatical nominal classes (i.e. differentiation of objects: big, small, round, flat, etc.), duality and triality (i.e. the primitive system of counting: two, three, many),—but also the languages with more abstract grammar. It is important, that the category of abstract plurality in a language (like later the notion of abstract set in mathematics) had a very late development.* The imperfection of the grammar of natural languages and its principal distinction from the logical syntax of the language of science may be seen in the absence of grammatical methods for the unambiguous expression of the principal abstract ideas, and also, in as numerous restrictions laid upon the formal grammatical systems. The grammar of natural language may be considered as an interpretation of a formal system. During the construction of a meta-language, it is necessary to translate the restrictions laid upon a grammatical system into lexics (i.e., to translate several grammatical meanings of natural languages by semantic factors of an abstract language, e.g., using the possibility of the expression of aspect meanings in natural languages by adverbs, meanings of cases by prepositions and postpositions, meanings of plural forms by a special word "many, much" etc.)†. So the formal system may be constructed on the basis of an interpreted one.

The logical syntax of abstract language (or meta-language for natural languages) should express only certain most general relational concepts. Consequently, the basic purpose of automatic gram-

---

* See on the Connection of Language and Mathematics: J. Lohmann, "Wort and Zahl," "Zeitschrift fur slavische Philogie," Bd. XXV, 1956, Heft I, p. 151-158; See also on the theory of "concrete sets": S. D. Katznelson "Historico-grammatic investigations," Moscow-Leningrad, 1949.

† See in the Vietnamese language the expression of relations by preposition and verb.

matical analysis is to formalize the relationships between the elements of the text with an indication of the sequence of operations to be accomplished, in order to determine an underlying logical structure.** The analysis of the text of language on the basis of formalization of the relations between elements is expected to be that corresponding most nearly to the principles of modern linguistics, that stresses the importance of relations in language (see the syntactic theories of Hjelmslev, their development by Bazell, Garvin, etc.).

In principle, it is possible to build up a field of syntactic relations which are indifferent to forms, carrying the corresponding functions in the real text. The syntagmatic function of a given form is defined by its position in the system of these relations. To characterize a unit of language, its position should be indicated in two fields: in the paradigmatic, and syntagmatic fields. For the discovery of basic syntagmatic relations a method of dissection of sentences on the syntagmatic level may be suggested; for each concrete language, a definite sequence of such syntagmatic analyses may be established. The characteristics of relations allow the identification and discrimination of the elements in the text, which often could not be done on the basis of pure distributive analysis, not utilizing the concept of syntactic relations. The paradigmatic content (i.e., the set of grammatic meanings) of a language unit is determined by its distributive properties and by relations of a syntagmatic nature. The recognition of a word may be accomplished by presenting simultaneously its lexical coordinates and those which define its relations in the text. The syntactic relations which are clarified by syntactic analysis, may be mapped into abstract machine languages,* where they may be expressed by single meaning logico-mathematical formulae.

One of the most important differences between the grammar of abstract language and the grammar of natural language is the fact that for abstract language all grammatical and lexical elements which connect the message transmitted to the author (and to the addressee of the message) are redundant. In other words, in abstract language the pragmatic plane is absent (provided that modality exists). The elements, relating a message transmitted to its author were investigated recently by R. O. Jakobson* and by E. Benveniste†. It was

---

** In a general form this suggestion about the goals of machine translation was expressed by P. P. Troyansky as early as in 1933, see P. P. Troyansky's "Translation Machine," Moscow, 1959.

* The correctness of this thought has been proved by the brilliant investigation of I. A. Mel'chuk concerning the interlingua, which is considered by him as a net of relations between the syntagmas of natural languages. See I. Mel'chuk "A model of interlingua for machine translation," "Abstracts of the conference on machine translation," Moscow, 1958, pp. 20-22.

*R. Jakobson, "Shifters, Verbal Categories and the Russian Verb," Russian Language Project, Department of Slavic Languages and Lit-
(footnote continued)

shown in their works that the system of verb and pronoun is built basically on pragmatic relations, determined by the structure of the act of communication (i.e., by the existence of speaker and listener). The most important difference between abstract language and natural one is the different aspect of communication which is characteristic of the language of science in general. Not by accident do the verbal categories in the scientific-technical texts possess specific characteristics, e.g., the use of impersonal and indefinite-personal constructions, specific use of the present tense in nomological statements (i.e., in those concerned with scientific laws); cf. also the peculiarities in using personal pronouns of the first person in scientific-technical texts. These peculiarities (explained by the absence of the regular situation of a two sided communication which left its landmark on the grammar of scientific-technical texts**) give to the grammar of abstract language its characteristic shape. This explains the absence of categories of person and other verbal categories which are called shifters by R. O. Jakobson, and also the absence of personal pronouns. In contrast to the usual point of view, the personal pronouns do not depend upon the context, but, on the contrary, they themselves determine the pragmatic value of the text, functioning as operators (quantors). In abstract language where the nonpersonal sum of all-the-humanity knowledge is recorded, expressions such as "my opinion is...," and "I discovered that..." find no place. It is obvious that the translation of similar expressions would only complicate one of the basic duties of the information machine, the determination between information which is new (interesting) and that which does not contain information, because it only repeats facts already known. The justified authorship is recorded in the information machine, not by pragmatic methods but by a semantic method (as a relation of a statement and its author).

5. As has been already indicated in the scientific literature,* it is expedient to use the abstract machine language of the information machine as an interlingua during machine translations from n languages to n languages. This may lead not only to the reduction of the number of algorithms (which used to be considered regularly while speaking on the subject), but also to the simplification of these algorithms (which is not considered regularly). Into the machine interlingua, the lexical and grammatical meanings of words and of the combinations of words of each of the n languages are mapped, however, in this case, the number of elementary units of vocabulary and the number of syntactic rules, determining the combinations of these units, would be strictly limited (because the meta-language is built

---

eratures, Harvard University, 1957. See V. V. Ivanov, "Code and Message," Bulletin, no. 5, N. 1957, pp. 48-50.

† E. Benveniste, "La Nature des Pronoms," Festschrift "For Roman Jakobson," The Hague, 1956, p. 36.

** W. V. Quine, "The Scope and Language of Science," The British Journal for the Philosophy of Science, v. 111, May 1957, N27, p. 8.

* Perry, Kent, Berry, op. cit., p. 130-131.

on the basis of splitting of the plane of content of each natural language into elementary discrete units). As the investigations of the linguists of the twentieth century (particularly, Sapir's, Whorf's, and others) indicate, the lexical systems of different languages do not correspond completely and therefore cannot themselves be immediately brought together. Without the construction of an interlingua, the problem of automatic lexical translation for the nonterminological lexics can hardly be solved. The representation of the semantics of each of the given languages by means of the set of the same elements, enables the establishment of a one-to-one correspondence between the words of different languages to be made. The advantages of translation by an interlingua is even more obvious when considering the grammatical analysis. The interlingua allows the avoidance of the difficult two sided grammatical procedure, the various stages of which can seldom be put into one-to-one relation (because the normal sequence of syntagmatic analysis of one language cannot be equated to the normal sequence of syntagmatic synthesis of the other language; for example, a Russian text cannot be constructed by the stages of analysis of an English text, etc.). As long as a part of the natural language text is redundant, the information machine gives a translation-abstract and not a literal translation.

6. The analysis of a text by machine may be considered as decoding; this point of view was expressed by Weaver in his first work concerning machine translation. Each time the machine is fed material for an analysis, the problem of the black box must be solved. It seems to be possible to design a machine, which is capable of building up an algorithm, for the translation of special texts from a n+1 language, on the grounds of the algorithms of translations from n languages. In the memory of such a machine, the semantic paradigmatic, and syntagmatic fields should be recorded (i.e. those possible combinations of meanings, which are present in the specialized text to be translated.) For example, the automatic decoding of uniform texts for a narrow field, where the syntagmatic conditioning of meanings is very great, represents a practical interest. For this purpose it is correct to build up an algorithm, permitting the recognition of ideographic signs or ideograms (for example, numbers) and of key signs (determinatives, for example, Mr. before the name, where Mr. is used for the indication that the name of person will follow.) The general features of the text and individual words may be determined by the ideographic and key signs; after this operation, the machine utilizes a statistical treatment of the text and combinatory substitutions for the decoding of individual words. The machine can itself determine the exact content of the text and construct the algorithm for translation of this text into abstract language on the grounds of the data indicated alone. The isomorphism of the decoded language and the already known languages, the existence of overlapping portions of the language to be decoded and of the languages known by the machine (i.e. the common lexical pool) may also be used for the composition of the new algorithm. For the construction of the theory of decoding, probably the ideas of the theory of games can be used.

Apparatus which would allow the machine to accustom itself to a given code (new to the machine) is also of extreme importance in solving the problem of oral translation, and consequently that of oral conversation of humans with the machine. In contrast to the acoustical apparatuses, the sound apparatus of a human is a self organizing one, which regulates itself depending upon the code to be percepted. Consequently, the following apparatus is necessary in order to enable the machine to perceive any human speech (not only one dictor's speech). The system of distinctive features of phonemes of the language should be programmed. The speech of every person, speaking the given language, is such a transformation of this system, at which its structure remains invariant. The machine itself must construct an algorithm, establishing a set of relations between the given individual code and the programmed phonological system. This problem is equivalent in several features to the problems of decoding, because the role of abstract language (i.e., the system of basic units, to which the given code is related) pertains to the phonological meta-language. On the grounds of probability determination of the phonetic norm, the most probable variants of a phonetic system of the language for a given time and place may be also programmed; this simplifies the construction of the algorithm by the machine, for the analysis of statistically rare variants.

7. The machine language, in the information-logical machine, is used for the further rearrangement of information, coded in the language given. Such rearrangement is accomplished by certain operations with the units of this language given. In the process of selecting information-containing statements, the machine itself can widen and change the organization of the language (vocabulary and rules). The interrelation between the natural languages and abstract language may be considered as the modeling of the relation of the language and of the thought in the information machine. However, for the modeling of the real relation between language and the thinking, the apparatus discussed so far is not satisfactory. For modeling the relationship of language and thinking, it is necessary to introduce data on experience (for example, by connecting instruments to the machine). Moreover, when information expressed in a language was considered above an assumption was made that it may be defined by a simple indication of the possibility of coincidence of the given word or text with a certain situation. However, natural language is characterized by the probabilistic character of the word meanings; that is, every word may relate a certain object, only with a certain probability. (The modeling of this relationship is possible in an information machine with stochastic elements which represent a type of an intelligence amplifier— described by Ashby*. This machine correlates information-con-

---

* W. Ross Ashby, "Design for an Intelligence Amplifier," "Automata studies," Princeton, 1956, pp. 215-234. In MacKay's article, printed in this collection, a scheme of machines which are able, besides the production of concepts, to perform meta-linguistic operations is de-
(footnote continued)

taining statements with the experimental data. Such a machine could be used for modeling the process of learning a language, provided that the machine extensively uses the method of Monte Carlo. Another problem is that of the immediate correlation of experimental data with the abstract language (the thinking of the machine). The problem of the development of an abstract language may be partly solved by the machine itself in the future.

A great theoretical and practical interest is represented in the modeling of disturbances of the relationship of language and thinking and the experimental data (for example, the disturbances in schizophrenia, corresponding to the pre-concept period in the development of a child, consist of the establishment of a connection between the syntagmatic system of natural language and the experience without the means of abstract language).*

The development of mobile system of speech while the stabile system of language is maintained unchanged†—(see the analogical interrelations between a stabile system and a mobile system which is based on the stabile one established for the genetics and the theory of evolution in the works of Ashby and other modern scientists)—pertains also to the number of language processes which may be modeled by means of machines. The mixture of languages (codes) and the transformation of one code into another (i.e., the historical development of language and the development of the families of languages) either by a given or by arbitrary algorithms can also be modeled by these machines. The transformation of a language by the algorithms of the general phonetic rules may be used for the prediction of the ways of linguistic development; in addition to this, the modern electro-acoustical apparatus may be also used for such experiments.**

8. The most important theoretical problem, put before linguistics and some border sciences by the design of an information machine and a machine interlingua for automatic translation, is the construc-

---

scribed: D. M. MacKay, "The Epistemological Problems of Automata," the same source, p. 235-251.

* See L. S. Vygotsky, "The Violation of Concepts During Schizophrenia," "Selected psychological works," M. 1956 (See also the "Thinking and Speech" in the same volume); H. Werner "On the development of Word Meaning," Cybernetics, "Transactions of the 7th Conference," New York, 1951, pp. 187-214.

† See the mistakes of Moget (МОГЕТ) type in an experimental translation from French to Russian by the machine "Strela," where, in the act of machine speech, the given units of language (morphemes) are incorrectly combined. Cf. V. V. Ivanov in "Proceedings of the Eighth International Congress of Linguists," Oslo, 1958, pp. 534-535.

** The author had the chance to be convinced of the experimental possibility of investigation of the process of the consonantal shift by means of an apparatus, which automatically changed the characteristics of stops.

tion of the general theory of codes (sign systems) and the construction of the theories of the interrelations between the codes. In the recent stage of modern linguistics, it has been found most advantageous to solve this problem for natural languages on the grounds of comparative linguistics. In this field, the following types of codes and of interrelations between the codes are established: the equality of codes, the intersection of codes, the isomorphism of codes, and the topological equivalency of codes. The parts of two languages of the same family are defined as two representations of the same abstract system—common (proto-)-language; they are simply isomorphic, i.e., they may be put in one-to-one-correspondence, which preserves the relations.* In the future, probably a practical conclusion can be obtained for these theoretical ideas and the practical solution of the problem concerning the automatic etymologic dictionary of related languages can be considered. In this machine, the algorithm of phonetic correspondence between related languages is to be preliminarily programmed. Words between which such phonetic correspondences are found are to be accepted as related, if, in their meanings, at least one semantic factor can be found which belongs to the same semantic field (or to one class of semantic fields). The source word is defined for the abstract system on the basis of correspondence of related words in languages—representations of one abstract system. All languages of the given family are run through the machine; however, only the related words (i.e., the intersection of vocabularies of isomorphic codes) are put in the permanent storages, while the words which form the differences within vocabularies of isomorphic codes, may not be recorded in the permanent memory, but be directed for treatment in the machine-etymological vocabulary of a language league (Sprachbund of non-related languages).

9. The method suggested above for the construction of an abstract language is immediately related to the construction of the theory of the interrelations between codes, since the abstract language is obtained through a definite transformation of natural languages. It seems possible to raise the problem concerning the consideration of the system of science as a language or code (or sign system). In connection with this, problems arise concerning the algorithmic transformations of the system or of the language of science, and the possibility of deductive derivations of one code - of one scientific system from another (when the accidental transformation and the accidental intersection may be used for the search of new possibilities

---

* V. V. Ivanov, "About Certain Concepts of Comparative Historical Linguistics," "Bulletin," no. 5, 1957, p. 53-54; V. V. Ivanov, "A Probabilistic Determination of Linguistic Time," "Problems of the Statistics of the Speech," Leningrad, 1958, pp. 66-67; V. V. Ivanov in "Proceedings of the Eighth International Congress of Linguists," Oslo, 1958, p. 26; V. V. Ivanov in "Abstracts of the conference on mathematical linguistics," Leningrad, 1959, pp. 6-7.

in the science). In the case of the information machine, the isomorphism of the two sciences or the two codes has a practical value because the system of abstract language, constructed for one science, may be used for the other one also, provided that isomorphism exists between them. In addition to this, the problems of isomorphism of the systems of the same science for different periods is important for the solution of the problem concerning the permanency (invariableness) of the system, developed for the information machine, e.g., the non-isomorphism of classical physics and modern quantum mechanics and the isomorphism of classical scientific genetics, the genetics in the interpretation of Schrödinger and the genetics in the works of Crick, Watson and Gamow. The last example is illustrative for the study of the interrelation of the isomorphism and the intersection of the codes: the theory, conceived in Schrödinger's book "What is life," is based on the intersection of the codes of genetics and quantum mechanics and the isomorphism of these codes; in the same manner Gamow's theory is based on the isomorphism and on the intersection of the codes of genetics and the information theory.

10. The modern ideas of cybernetics, which are based on a partial intersection and isomorphism of sciences, arose as a protest against the excessive specialization, thanks to which every specialist learned "more and more about his always less and less." This protest in the theory led to important practical results. In particular, the construction of information machines may solve in a satisfactory degree the practical problem surrounding excessive specialization, and release from the memories of scientists a load of unnecessary knowledge. In many articles and books printed recently the possible consequences following the design of information machines are compared with those which followed the invention of printing. At the end of an article devoted to the linguistic problems of the construction of information machines, it is correct to emphasize that these machines may cause a far reaching influence on the development of the language, (such as printing of books caused in its time). At present language has been influenced by the methods by which information is transmitted: through time (books), and through space (by systems of telecommunication). The information machines may unite both of these methods of transmission, because of Wiener's definition "a pattern is a message and it may be transmitted as a message."* This is a perfectly realistic point of view when the information machines are considered, since information machines may transfer the patterns (codes) put into them or received in them along the channels of telecommunication. Due to the great significance of these dynamic libraries of the future, it is necessary to concentrate the united forces of a great number of our linguists working in close contact with the specialists of other sciences in order to construct such systems.

---

* N. Wiener, "The Human Use of Human Beings," New York, 1956, p. 96.