[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 49

The Universal Code of Science and Machine Languages

N. D. ANDREYEV

Experimental Laboratory of Machine Translation, Leningrad, U.S.S.R.

The development of cybernetics presented the problem of machine utilization of linguistic information in its fullest extent. The associated problems represent today the most important section of applied and mathematical linguistics; moreover, they begin to exert a far reaching influence upon the development of linguistics as a whole.*

I would like to draw attention to the connection existing between the utilization of linguistic information in machines and the standardization of scientific language.

Linguistic information, entering the machine, has either a literal or a phonetic code representation at the input. The transformation from the latter to a representation in the form of the most widespread binary or any other machine code does not change anything in the structure of the message†. It is, therefore, merely a case of code transformation, similar, for example, to the transition from optical signalization to radio telegraph transmission.

The process of translating a message from an input to an output language (or, within the framework of the same language, the transposition of a message from one style to another) is different in principle from the above. The invariant element here is the content of the message; while its structure is a free parameter, insofar as the input and output languages are, as a rule, not identical,** a variation

* General discussion of the modern developmental tendencies of our science, related to the new equipment, is presented in the article, "Basic Problems of Applied Linguistics" by N. D. Andreyev and L. R. Zinder (Problems of Linguistics, 1959 No. 4).

† I am deliberately avoiding the questions of rhythm, intonation, and individual peculiarities of speech, that render the literal representation non-equivalent to phonetic representation. In principle, the machine may also contain additional information conveyed by the above elements of sound speech.

**A quasi-identity of the input and output languages may occur, for example, in an editing machine, or in a machine which automatically widens its translating algorithm by accumulating experience.

of this parameter is unavoidable. Such a variation is, indeed, the essence of machine translation which can be rightly called the transformation of the structure of the message, or, briefly, message transformation.

Both of the above processes substantially differ from such operations as abstracting (reviewing) of text or compiling information obtained from various sources. In those cases, not only the formal structure, but also the content of the message cease to be invariant, and are transformed into a partially or fully variable parameter. In the abstracting of a text, three operations are performed: 1. Determination of the meaning of the message; 2. Evaluation of its significance; 3. Acceptance or rejection of the message (or of its part) according to given criteria. As a result, the meaning content of the output information is reduced. When information is compiled from various sources, a fourth operation is added to the other three: the consolidation of messages. Here also the total meaning volume of the output information at least does not exceed the total input volume.

Operations of this type may well be termed the processing of the message content or briefly message processing, to distinguish them from the transformation of message structure.

The operation of searching of messages possessing a given content is a special case of message processing, although this is not self evident. Indeed, such an operation again has the elements: 1. Determination of meaning of the surveyed messages; 2. Evaluation for correspondence with the given content; 3. Acceptance or rejection in accordance with the results of the evaluation. If acceptance is carried out in the form of duplication, the volume of the duplicated information will, in the general case, be lower than the volume of the surveyed information; if acceptance is followed by the extraction of information, the same will apply to the extracted information (and also to the surveyed information). In both cases, the content of the input information flow has been changed, i.e., the message has been processed. It may be said that searching of information with a given content means abstracting under specific and particularly rigid criteria of selection.

According to the various types of utilization of linguistic information in machines, various machine languages are being developed.

The first stage of message transformation consists in a formal analysis of input text in order to determine its structure; the second stage is the synthesis of the output text according to the determined structure (which may be varied within the limits of invariance of the message content). These two stages may be directly coupled (the binary method) or may be performed without reference to each other (the system of independent analysis and independent synthesis). In both methods, the transformation algorithm should be applied to information of three types: the text in an extra-machine language (i.e., input or output language); information on text structure obtained in the course of analysis or given before synthesis; and information on the structure of the transformation algorithm. The three classes form the para-language, meta-language, and ortho-language levels of information; the totality of the last two represent the meta-language of information transformation (i.e., machine translation or machine editing). We shall call it machine language of the first type (ML-I) .

Every language is based upon a set of symbolic units and systematic relationship among the units, is determined by probability characteristics of its structural elements and strata, and is characterized by sets of operators generating the messages.*

None of these aspects may be absent in a language designed for communication; machine languages are no exception to this rule.

The ML-I system is formed by its meta-language or ortho-language elements (para-language elements do not belong to ML-I by definition). The probability characteristics of ML-I represent the function of two arguments: typology of the processed extra-machine languages and the structure of transformation algorithms. The fact that the formal aspect of these algorithms is far from fully determined by the typology of the treated language is clear if only by virtue of the possibility of forming algorithmic abbreviations. In general, the structure of algorithms depends to a large extent upon the methods of representing information.*

Since a message in the ML-I language is represented by a definite instruction for searching or transformation of information, the set of operators generating such a message comprises an operational syntax of algorithm formation. It is clear that systems of elementary operators of this type, being developed by O. S. Kulagina (Moscow) or by V. Yngve (Cambridge, USA), converting the notation of the algorithm into a program directly executed by the machine, and creating the basis for automatic programming, play the role of a sub-language with respect to the ML-I. †

When the transformation of information is accomplished by methods of independent analysis and independent synthesis, an intermediate language is a mandatory and, essentially, the central component of the system. The independent analysis is carried out as a transition from the input language to the intermediate language; while the independent synthesis is a transition from the intermediate to the output language. The intermediate language shall be called machine language of the second type (ML-II).

A number of approaches toward ML-II are possible.** The ML-II construction method largely depends upon its purpose: If ML-II is to

---

*See Andreyev's article on "Algorithmic Modelling of Language Based Upon Statistical and Combinative Structure of Speech" (Materials on Mathematical Linguistics and Machine Translation, vol. 2, 1959).

*See articles by B. M. Leykina & S. Y. Fitialov and O. B. Frolova & S. Y. Fitialov in the second collection: "Materials on Mathematical Linguistics and Machine Translation."

† One can draw an analogy between the translation from ML-I to the operational sub-language and the substitution of literary style by

(footnotes continued on page 1064)

be used only within machine translation, its structure may be totally determined by the aggregate of languages subject to translation; on the other hand, if the message transformed into ML-II is intended to emerge beyond the translation machine, the ML-II language should be autonomous, i.e., it should have its own structure of messages, independent of the particular extra-machine language from which they were obtained.

It is quite clear that autonomous ML-II, relieved from an excessively rigid association with extra-machine languages, is most convenient in a form convertible to other machine languages (for example, to the informational language). Moreover, autonomous ML-II, because of its maximum economy in comparison to extra-machine languages (and to ML-II languages of the correlation type which fully retain and even accumulate the excess elements of extra-machine languages), is most convenient in the form of an intermediate information storage. Sooner or later, an international system of machine translation will be created, whereby each national center will carry out translations from the local language to ML-II and will duplicate the obtained perforated tapes. The copies of messages encoded in ML-II will be distributed among the other national centers where the tapes will be fed directly to the machines for translation into the appropriate local output language. It is readily apparent that with such a system each local center will require only two algorithms of transformation: from the national language to ML-II and from ML-II to the national language.

Still more important is the possibility of transforming information generated by automatic devices into messages in ML-II; the information in this form will be equally suitable for any zone of the planet. The messages may be received simultaneously in all zones with an automatic local decoding into the language of the zone.

In this manner, an autonomous ML-II may find a much wider field of application than a language of the correlation type which represents merely a network of relationships between the input and output languages of machine translation.

Since the ML-II should by definition have a permanent contact with extra-machine languages, its structure may not be fully logical: a strictly logical structure of such a language would lead to extremely cumbersome algorithms of message transformation. The requirement of maximum simplicity of the transformation algorithm lies at the base of the idea of the intermediate language. The optimum characteristic of the ML-II will be achieved by retaining the invariance of scientific meaning of a message during its transformation, on the one

---

a professional jargon (if one is not afraid of hurting the feelings of the machine).

**See general evaluation of these approaches in the report by N. D. Andreyev and S. Y. Fitialov, "Intermediate Language of Machine Translation and Principles of its Construction" (Theses of the Conference on Mathematical Linguistics, L., 1959).

hand, and by performing such transformations in the simplest and fastest manner, on the other.

The set of symbolic units and the system of relationships in the ML-II is determined by classifying the grammatical information of input languages into nontautological, tautological, and sub-information.* The probability parameters of ML-II elements are connected with the probability characteristics of the most frequently used extra-machine languages by the requirement of optimum congruency,| which insures the highest simplicity and rapidity of transformation for the entire field as a whole. The operators generating messages in ML-II, used only for machine translation, are to a considerable extent determined by the properties of the area of extra-machine languages involved. In the ML-II language used beyond machine translation, the set of generating operators is independent of these properties.

The ML-I and ML-II languages deal with the transformation of messages without regard to the meaning of the latter. On the other hand, machine languages intended for the processing of messages, are linked with the meaning analysis of text and are, therefore, of a somewhat different form.

The analysis of textual content by an algorithm deals with information of four, rather than three, types: along with information on the structure of the message, there is information on the semantics of the message, partly based on the structure information and partly derived from other sources. In other words, the meta-language level of information is split into a meta-grammatical (dealing with formo- and tectoglyphy) and meta-semantic (dealing with semoglyphy) levels; the center of gravity is shifted to the analysis and processing of the semoglyphy of the input text.**

A hierarchy of levels, more complex than in ML-I, leads to a new code, the meta-language of information processing (that is, machine abstracting, logical analysis and selection in informational machines, and searching in an automatic reference service). Such a language shall be called machine language of the third type (ML-III). The system of ML-III is formed by its meta-grammatical, meta-semantic. and ortho-linguistic elements. Since the semantic analysis of text, retaining a constant and fairly strong association with formal analysis, is basically logical, it follows that the probability characteristics of

---

* A more detailed treatment of this concept is contained in the article by B. M. Leykina, "Two Types of Grammatical Information in their Relationship to the Intermediate Language" (Materials on Mathematical Linguistics and Machine Translation, vol. 2, L., 1960).

† See article of N. D. Andreyev "Machine Translation and the Intermediate Language Problem" (Problems of Linguistics, 1957 No. 7).

**For a discussion on the classes of language hieroglyphy (formo-, tecto-, and semo-glyphy), see article by N. D. Andreyev, "Meta-Language of Machine Translation and its Application" (Materials on Machine Translation, Vol. 1, L., 1958).

ML-III elements represent a function of three arguments: the selected logical system, semantic typology of the processed extra-machine languages, and the structure of algorithms of information transformation. A communication in ML-III is the instruction to search information with a given content or to process it according to established criteria. Accordingly, the set of ML-III operators comprises an operational syntax whose hierarchy is more complex than in ML-I by one or several degrees.

In the transformation of information, the algorithms are fixed by the agency of ML-I, while the transformed message is reduced to a sequence of symbols in ML-II. In a similar manner, the processing of a message consists in representing the algorithm by ML-III, while the processed message itself is converted by this algorithm into a special sequence of symbols which shall be called machine language of the fourth type (ML-IV). As an example of an ML-IV language may serve the informational language, i.e., a code which records and stores the accumulated information in the information machine.* It is clear that the input of text fed into such a machine should have been first converted into ML-II, rather than directly using the extra-machine language, if only for the sake of having a single algorithm of semantic analysis in the machine (on the other hand, if the input were limited to a single extra-machine language, as it is intended by a number of systems under development, it would definitely reduce the range and value of accumulated information). In addition, it can be readily seen that the low waste and high orderliness of the ML-II language, as compared to extra-machine languages, render the transition from ML-II to ML-IV somewhat more simple and more accurate from the point of view of the algorithm, than transitions from extra-machine languages directly to ML-IV.

According to the same considerations of linguistic universality and algorithmic effectiveness, the extraction of cumulated information from the machine should preferably have the form of a reverse transition from ML-IV to ML-II. In the future, the development of informational service will not be based on the establishment of dozens or hundreds of local cumulative centers, each of which would inefficiently duplicate the others' work, but will consist in expanding branch information storage centers specializing in definite areas of science or technology. Consequently, an efficient organization of an international network of information machines should be based on the extraction of information from the machines in a form convenient for a universal, and not narrowly local, utilization. Finally, the network of information machines, built according to the branch principle *is* found to be naturally associated, in the process of converting information from ML-IV to ML-II, with the system of machine translation centers

* See reports by G. E. Vleduts & V. K. Finn and N. M. Yermolayeva & E. V. Paducheva, delivered to the Conference on Mathematical Linguistics (Theses of the Conference of Mathematical Linguistics, 1959).

described earlier in this report. Just as in the case of machine translation centers, where a pair of algorithms is sufficient (from the national language to ML-II and back), the specialized information centers will also need only two algorithms: from ML-II to the branch sub-language of ML-IV and back.

If the autonomous nature of ML-II structure is a functionally variable parameter (see above), in the case of the ML-IV language, its autonomous nature is a mandatory characteristic. This is clear in relation to informational languages designed for storage of information in machines; abstracting is also associated with the production of a definite text. It may be objected here that machine abstracting (reviewing) of single-language texts may have an output in the same extra-machine language that has been used at the input stage; however, it may be noted that even here the machine may operate with a pair consisting of ML-III plus ML-IV; at the input, the extra-machine language is translated into ML-IV (which is essential for the analysis of meaning) while the output may result from the reverse translation. Thus the information will be, although temporarily, encoded in ML-IV. i.e., the autonomous characteristic is unavoidable even in this case.

Speaking of ML-IV, one should point out the fact that in the majority of cases it will not be used in its entirety, but in the form of branch sub-languages (examples: informational language for chemistry, abstract language for cybernetics, code language for patent searching, etc.). Within those languages, the entire complex of factors such as the system of elements, the probability characteristics, and even the hierarchy of operators, will be mainly determined by the pragmatic circumstances of the area of knowledge involved. Nevertheless, in spite of the diversity of aspects of these sub-languages, they may and will be generalized into an ML-IV language of a general type, particularly because of the many convincing proofs of the importance of <u>revealing the links</u> among data from various sciences. Such an exposition of interdisciplinary links will be still more important in the future and will successfully involve multi-branch informational machines for this purpose.*

It is therefore necessary to foresee a parallel utilization of ML-IV in the form of branch sub-languages and in a general multi-branch form (it may also be possible that such a parallelism will prove convenient in the utilization of ML-II).

So far we have discussed languages used to analyze and synthesize messages <u>in a given code</u>. The development of research during the past few years in the modelling of language structure has shown the feasibility of algorithms† capable of solving the problem of determining the linguistic code by analyzing a given aggregate of messages.

---

*The author notes with satisfaction that a similar thought (formulated even more broadly) has been advanced by V. V. Ivanov in his remarkable paper "Theoretical and Applied Linguistics" (Materials on Mathematical Linguistics and Machine Translation, Vol. 2, L., 1960).

(footnote continued)

Moreover, it was possible to construct algorithms determining certain characteristics of semantic system either within the framework of a single language or through a multi-language translation.**

The machine realization of such algorithms, modelling the language according to the speech data, requires a special meta-language of linguistic analysis which shall be called the machine language of the fifth type (ML-V). Similarly to ML-I and ML-III, ML-V has meta-language and ortho-language levels although the latter is in this case substantially more complex, since the language-modelling algorithms contain a large number of computing operations whose results determine the selection of direction during conditional transitions. The probability characteristics of ML-V, in a sufficiently wide application, will approach the statistics of the meta-language of general linguistics. The ML-V set of operators differs from the corresponding sets of the ML-I and ML-II languages mainly by the presence of a special group of operators converting numerical data into linguistic information. It may be said that this portion of the set of ML-V operators transforms the linguistic information from an implicit into an explicit form. Insofar as the determination of code structure from the messages represents a special case of a more general problem of deducing the structure of a system from its manifestations, one can predict that ML-V and its associated algorithms will be used beyond the limits of linguistics (for example, to decode genetic codes on the basis of data of phylo- and onto-genesis or to trace a system of neuron connections in the brain from its external manifestations).

The juxtaposition of machine languages within the algorithmic group (ML-I, ML-III, ML-V) and an analogous comparison of communication machine languages (ML-II and ML-IV) as well as a comparative analysis of both groups, constitute the basis of machine language typology which at this time has earned the right to be termed the experimental typology of languages.

It was noted above that the relationships within the hieroglyphy of the ML-I and ML-III languages are not identical: the center of gravity for the ML-I is the analysis of formo- and tectoglyphs, whereas the study of semoglyphs is the basis for ML-III. The boundaries among the three areas are determined not only by distributive but also by probability factors. The differentiation of the three areas of hieroglyphy in the construction of any type of machine language is regarded as a problem of finding the optimum functions of systematic pattern

† See the cited work of Andreyev on the algorithmic modelling of language from speech data in the same 2nd vol. of "Materials on Mathematical Linguistics and Machine Translation."

**See article of Andreyev "Modelling of Semantics as a Quasi-Normalized Space" (Materials on Mathematical Linguistics and Machine Translation, Vol. 2, L., 1960).

See the cited report of Andreyev and Fitialov on the intermediate language delivered at the Conference on Mathematical Linguistics.

and frequency according to the ultimate purpose of the machine language. This may be exemplified by the different treatment of category information for two different types of machine languages. The language categories can be regarded as certain variables within a given finite range of values. For example, let T be the variable of grammatical tense, where T_1—past tense, T_2—present tense, T_3—future tense, T_0—no definite information on tense (indeterminate tense which should not be confused with the absence of the tense category itself). The range of values of the variable T will then represent a conjunction (T_1 and T_2 and T_3) which shall be termed common tense: "The earth revolved, revolves, and will revolve around the sun." On the contrary, an indeterminate tense is disjunction of values of the variable T: (T_1 or T_2 or T_3). Therefore, the common category characteristic differs from the indeterminate characteristic just as the full range of values of the variable differs from its partial values. With regard to the representation of these two concepts, we might note that the selection of their designation form is not determined by distributive considerations but by probability considerations. More widespread common tense should be designated in both the ML-II and ML-IV languages by the absence of the formoglyph (functional zero), while the indeterminate tense, encountered fairly rarely in extra-machine languages, should be designated in ML-II simply by a disjunction of all the elements of the tense variable values. In the ML-IV language, where the meaning analysis will reveal it somewhat more frequently, but not frequently enough for the formoglyphic representation, it may be designated by a special semoglyph.

Regardless of the differences in the parameters of machine languages, they have many more common than specific elements. Therefore, they should be always regarded as branches of a certain universal machine language (ML-Σ). Such an approach pursues two aims: first, it facilitates the interaction between various machine languages and, consequently, the utilization of each language (examples have been given above); second, it makes possible to coordinate machine languages with extra-machine character systems, both symbolic and linguistic. Much has already been said and written about the extra machine character systems and their unification. Therefore, this paper shall be limited to as brief an exposition of the problem as it is permitted by the basic topic.

An entire group of symbolic languages has been accumulated and developed in recent times: the symbolic language of mathematics studied by us in school (SL-M), the strictly formalized language of mathematical logic (SL-L) chemical symbolism (SL-C) and its offspring symbolism of nuclear physics (SL-NP), and the fairly peculiar character language of formal genetics (SL-FG); this list may also include the symbolism of linguistics (SL-LA), particularly after the appearance of laryngeal theory in the science of Indo-European languages and the development of structural analysis. Finally, we have witnessed the birth of the rapidly developing symbolic language of machine translation (SL-MT) whose foundation and application is presented in the work of the Experimental Laboratory of Machine Translation.*

It is known that the standardization of phraseology and its subsequent abbreviation is the beginning of any symbolism. A more profound study of the origin and evolution of symbolic languages shows that symbols are first developed for those concepts that are encountered more frequently than others. The use of symbols for one concept leads, as a rule, to the appearance of symbols standing for the associated concepts; this means that symbolism develops in a more or less systematic pattern: as in speech, the composition and organization of symbolic languages are determined by the probability, distributive, and algorithmic factors.

The development of symbolic languages is closely connected with the progress of scientific theories. Most theories (not excluding the mathematical theories) are born in a nonformalized or weakly formalized shape; they are given the necessary formalization only in step with their refinement and development. The formal apparatus of modern theories occurs primarily as computations, i.e., as aggregates of rigidly determined rules for symbol processing. The availability of such formal apparatus based on well developed symbolism provides the investigator with a very convenient instrument, facilitating his search and bringing to the surface many problems which would not be obvious without the help of symbolic language. Finally, another advantage of symbolic language shall be noted, of particular importance in the age of internationalization of science: symbolic language is invariant with respect to national languages. Therefore, there are enough reasons to develop symbolic language further in those sciences where they already exist and to create symbolic languages in the areas where they are still absent. Moreover, we are now witnessing the beginning of contacts among symbolic languages of different sciences: SL-L and SL-M have already begun interacting, the link between SL-C and SL-NP dates from the origin of nuclear physics, while the SL-MT has from the very beginning been built of elements from SL-L and SL-LA,

It is now time to make a basic study of the comparative typology of symbolic languages. This would lead not only to the creation of a general theory of symbolic languages and, therefore, to a better pattern of evolution of each language but, what is more important, to the ultimate unification of all symbolic languages into a universal symbolic language (SL-Σ), allowing for a uniform encoding of formal chapters of individual sciences and, what is most significant, for the use of symbolic notation in boundary areas of science where branch symbolisms will unavoidably begin to overlap one another. Just as in the case of ML-Σ, SL-Σ will have an internal function of coordinating the branch symbolic languages (sub-languages of the universal

*See "Materials of Mathematical Linguistics and Machine Translation" (from 1958) and "Reports and Communications on Mathematical Linguistics and Machine Translation" (from 1960).

symbolic language) and an external function of coordination with machine (ML-Σ) and linguistic character systems.

The most difficult standardization to achieve pertains to linguistic systems although even here considerable research work has been done. Language standardization is being carried out in two directions: the establishment of terminological systems in separate branches of knowledge, and the coordination of national terminological systems. The most effective method of bringing order to terminologies is a joint effort on the part of subject specialists and language specialists. An example of this planned attack on the evolution of a specialized subject language has been furnished by the chemists who have developed not only a system of chemical concepts and terms but also well founded methods of forming new terms. A terminological code may be said to exist in chemistry (TC-C), geology and mineralogy (TC-GM), medicine and pharmacology (TC-MP) and a nomenclature code in astronomy (TC-A), zoology (TC-Z), botany (TC-B), and in trade and service (TC-TS).

The work on coordinating national terminological systems is in a much greater need of planning. The multilingual dictionaries published by the International Commission on Electrical Engineering demonstrate one of the possible organizational methods; another method would be the developing practice of homotopic abstracting in foreign languages in periodicals and monographic publications.

It should be noted that the standardization of word languages of science is not only developing with respect to terminology and nomenclature; similar processes occur in the field of phraseology. It has been long noted that many turns of speech, becoming canonical, spread from one language into another thanks to translations, and from one area of knowledge into another across fringe areas of science (an example is the current adoption of mathematical phraseology by linguists working in the field of mathematical linguistics).

Naturally, the most important is the standardization of the meta-languages of science. At the same time, it is the most difficult, because the meta-language of each science is affected by the divergencies of the interpretation of data procured by the given science. This is the least likely area to depend upon custom and accidental convergence of terminological systems.

The theoretical development of terminological practice leaves much to be desired: we do not have a true fundamental theory of terminology construction. Neither will such a system be developed until the necessary spadework on the establishment of the typology of terminological systems has been completed. In this case, the comparative typology has two dimensions: a juxtaposing analysis is imposed upon systems of various sciences and systems of various languages. Such a basis may be used for the theoretical development and practical realization of a universal terminological code (TC-Σ), in relation to which specialized terminologies will play the role of sub-codes, and national systems of terms, that of positional realization (allocodes in terms of structural methodology).

We have approached a formal problem concerning the relationship among the universal machine language (ML-Σ), universal symbolic language (SL-Σ), and universal terminological code (TC-Σ). The sources of these three systems are fairly diverse and their paths of development are not uniform. It would be difficult to expect, therefore, any initial isomorphism among these systems. The conditions of their use are also of a fairly different nature: ML-Σ is used in machines, SL-Σ and TC-Σ outside of machines; on the other hand, ML-Σ and SL-Σ are invariant with respect to national languages, whereas TC-Σ involves a multiplicity of realizations. These difficulties, however, can and should be overcome. The need to coordinate the three Σ's is obvious; the feasibility of such a consolidation resides in the unique nature of symbols, common for ML, SL, and TC.

The invariant element (signeme)* of the machine, symbolic, and terminological notation is meaning (gnoseme) based on a certain unit of objective reality (realeme). Various manifestations of the objective unit (alloreals) have never prevented the talented student from discerning the common element of reality, arriving in his mind upon the appropriate meaning, and selecting the suitable linguistic symbol. The meanings derived by various researchers for the same realeme do not always coincide in all their aspects and practically every time appear as an aggregate of allognoses. Nevertheless, scientists still understand one another, proving that as a group they develop, sometimes spontaneously and sometimes deliberately, a certain invariant form of the aggregate. Such an invariant, based upon the realeme, is handed down to the next generation as a common property—the gnoseme. The further science progresses, the fewer allognoses are left of the inherited gnoseme (we all know that long established concepts are understood in practically the same way by the majority of researchers). This implies a natural tendency towards the derivation of an effective isomorphism between groups of realemes and groups of gnosemes.

The term developed for a concept in terminological code, the corresponding symbol created in symbolic language, and the machine character designed for this concept in machine language, can always be brought into a one-for-one correspondence. Conversely, the established one-for-one correspondence yields the character unit, or signeme, whose various manifestations, or allosigns, are the term, the symbol, and the element of machine language. There is no doubt that in the foreseeable future this triple representation of the signeme will remain a constant factor; a more distant future, however, will probably place humanity face-to-face with character systems of a completely different nature.

The study of the relationship among the given allosign, the given allognose, and the given alloreal may lead to fairly informative results in regard to the psychology of scientific creativity and the origin

*Translator's note: this, and a number of following words of Greek origin used by the author, have been retained in their original form.

of notational systems. However, this is not particularly significant for science as a whole: a general scientific significance should mainly be sought in the study of relationships among the given signeme, given gnoseme, and given realeme. The fixation of uniquely determined relationships of this kind is the necessary basis of the development of a universal code of science (CS-Σ), in relation to which, ML-Σ, SL-Σ, and TC-Σ, will play the role of three variant representations, ideally differing only in form and function, but not in the system of meanings. The road towards the creation of the universal code of science cannot but be long and arduous; however, since under all circumstances it will have to be traveled, it is best to form a clear picture of its nature. Two ingredients of CS-Σ, i.e., terminological systems and symbolic languages, although poorly developed, already exist. The creation of the third ingredient, machine languages, is only being attempted at this time. A proper understanding, at the very beginning, that this activity ultimately leads to the universal code of science will render the efforts of researchers working in this field more purposeful and effective.