

Voprosy Yazykoznanija #5, pp 37-49, 1976

LINGUISTICS AND THE CONTEMPORARY STATE OF MACHINE TRANSLATION IN THE USSR  
by R. G. Kotov

During recent years Machine Translation received considerable public attention in connection with analyzing solutions for further development of the State system for scientific-technical information (Gosudarstvennaja sistema nauchno-tekhnicheskoj informatsii). MT is treated as one of the components of this system. Within this framework MT is considered not as an exciting theoretical field, but rather as a practical tool for obtaining large amounts of translation of scientific texts of "rough" quality for purposes of information retrieval services.

During this period there began to appear reports in foreign countries describing successful and economically profitable application of the computers for massive commercial translations. However, this changing viewpoint was not noted in our country since the established opinion in the USSR was that MT is a task for the future, and that it will take a long time to work out the theoretical fundamentals of the theory of translation.

There was another fact which escaped the attention of the linguistic community in the USSR. Toward the end of 1973 and the beginning of 1974 a special temporary committee on science and technology was organized under the auspices of the State Committee of the Council of Ministries of the USSR (Gosudarstvennyj Komitet Soveta Ministrov SSR po nauke i tekhnike). This committee, composed of representatives of various organizations interested in practical MT, specialists on automatization of informational processes, specialists on MT took as its task determining under what conditions MT could be developed as a practical system at the present time. This should be a working, expedient MT system.

Under the "practical (working) MT system" is understood a system of dictionaries automatically used, equipped with necessary linguistic information and programs in order to produce massive "rough" quality of scientific technical texts. The editing of this MT translation should not take more effort than the editing of the usual translation. The "rough" quality means that the translated text is understood by the user in terms of clearly presented meaning; the meaning corresponds to the meaning of the source, and, therefore, this kind of translated text could be used as a source for information.

The conclusions this committee arrived at after a detailed analysis of the state of affairs in the MT field, both in the USSR and abroad, are the following:

"The level of achievements both in theories and experiments on MT makes it feasible to raise the question of moving toward the practical realization of MT in the USSR."

"The economic significance of practical MT could be evaluated on the basis of the following assumptions/assertions:

--MT processing of the text is approximately 5 times cheaper than human translation...;

--The time-consumption required by input of the text into the computer could be compared with human retyping of the text, translated by humans

--In terms of speed MT (including the post-editing) could be achieved at least ten times quicker;

"The work for creating practical MT and its deployment should be carried out already at present time, without demanding the preliminary conditions for solving all the theoretical problems aimed at producing translation of a higher quality."

The Committee has also stated that the absence of practical MT works became a hampering factor for further development of retrieval research in general, not just that of MT in particular.

The basic conclusion is that the USSR has no practical MT system despite the existing achievements in theory and practice and the real opportunities, while in foreign countries MT has entered an era of commercial application by both state and private organizations.

What are the causes of this state of affairs in the MT field in the USSR? A brief history of MT development within the USSR should be presented. Certain facts in its development might help in understanding the peculiarity of MT development in our country.

#### THE FIRST STAGE (1954-1958)

In the Institute of precise mechanics and computing technology (Institut tochnoj mekhaniki i vychislitel'noj tekhniki) of the Academy of Sciences of the USSR under the guidance of D. Yu. Panov and I. S. Mukhin and other members of their group (which included L. N. Korolev, S. N. Razumoskij, and the linguist I. K. Bel'skaja). The first Soviet experimental MT translation from English into Russian was made on the EVM BESM computer in December of 1955. Then, in 1955, the group under the guidance of A. A. Ljapunov and his assistant O. S. Kulagina carried out experiments from French into Russian using the EVM "STRELA" (the linguistic work was done by I. A. Mel'chuk and T. N. Moloshnaja). The results of these two groups were reported by the senior researchers of both groups in co-authored papers<sup>1</sup>

---

1. D. Yu. Panov, A. A. Ljapunov, I. S. Mukhin, Avtomatizatsija Perevoda s Odnogo Jazyka na Drugoj, M., 1956.

Without commenting on the questions connected with coding techniques on the computers, differences in using the linguistic information and designing the transfer algorithms could be reduced to the differences between "empirical" (I. K. Belskaja) and "analytical" (I. A. Mel'chuk, T. N. Moloshnaja) ways of solving the same problem. However, one has to note not so much the differences as the fact that the leaders of both groups recognized the existence and justification of different attitudes for solving new and complex MT problems.

In that very paper Panov warned about the danger of being carried away by logical analysis of the language structure as a tool for solving the MT problem. This logical way seemed to be attractive especially if one would follow the direction of certain MT work in the USA, particularly from the mathematical point of view since it made it possible to formulate the MT problem as essentially a mathematical problem. However, "the very nature of the translation is such that one can not completely ignore the individual features of the input text. Evidently, we encounter here a problem which requires special analytical methods, similar to those experimental methods which are used in studying natural phenomena."<sup>2</sup>

The group of D. Yu. Panov has also formulated the basic principles of designing a MT algorithm, some of which are valid at the present time, too:

- the maximal separation of the dictionary from the programs;
- storing in the dictionary of the inherent grammatical features of words;
- determining the meaning of the polysemic words on the basis of the contextual environment, their grammatical features and the analysis of the grammatical structure of the sentence, and other factors.

---

2. Ibid., p 15.

After publication of the results of these first works on MT there began emerging various groups of MT enthusiasts in our country who on the whole did not have the opportunity to test their algorithms on a computer and were engaged essentially in theoretical work.

#### THE SECOND STAGE (1958-1961)

The majority of participants in the first MT conference in the USSR in 1958 despite the warnings of Panov as mentioned above, found the direction of Bel'skaja too "empirical." The directions for research suggesting following some models in formal terms and using the intermediary language in order to recreate the logical structure of a natural language were recognized as more attractive and more promising. The discussions on the conference indicated not only the differences in opinions concerning the ways of going about solving MT problems, but also the inclination of some researchers to consider the direction of their research as the only correct one. The stage of suppression of "empiricists" and the start of fruitless searches for "universal solutions" of MT tasks was completed at the ALL UNION CONFERENCE ON INFORMATION PROCESSING, MACHINE TRANSLATION, AND AUTOMATIC READING OF TESTS IN 1961 (Vsesojuznaja konferencija po obrabotke informatsii, mashinnomu perevodu i avtomaticheskomy chteniju teksta v 1961 g.). At the final session it was announced that according to prevailing opinion MT as a practical problem should be removed from the agenda and all efforts should be devoted to working out the theoretical basis of translation.

The group of Panov, Mukhin and Bel'skaja ceased to exist at that time. The work of other groups came slowly to an end.

### THE THIRD STAGE (1961-1974)

This period is characterized by the prevailing development of theoretical studies of language outside of any connection with the specific task of designing practical working MT systems.

It is necessary to note that during the period of the late '50s and early '60s "reevaluation of values" in the field of MT has also taken place in foreign countries. The works carried out made it possible to come to certain important conclusions.

1. It turned out that the existing grammars and the experience in formalization of linguistic data were inadequate. There was no formal apparatus for describing morphology, syntax and semantics to the degree that they could be used in designing MT algorithms. This, in particular, served as a stimulus for development in various directions of structural and mathematical linguistics.
2. It became evident that one should test on the computer not only algorithms, but also theoretical constructs in linguistics, without which one can not evaluate their applied importance for MT.
3. Designing an MT system even in its simplest variety should not be considered as temporary work, but rather as a consuming long term task, whose success can be guaranteed only by simultaneous efforts of linguists, programmers and computer engineers.

The conclusion concerning insufficient access to computers and the resulting discussion that the special purpose computers should be built in the immediate future lost its relevance since the new computers with large memories were built, coupled with high speeds and mathematical operating systems.

However, most important was the awareness of the fact that language studies and the search for formalizing language structures are necessary not only for MT but also for more general procedures for treating the information by computers as well as in the interest of the development of the theory of language. However, while the language studies in MT terms are both defined and delimited by their applicability for computer testing, no such polar limits are assumed for the formal model studies concerning language theory or cybernetic problems in general.

Thus, there developed two trends in linguistic studies, applied and information retrieval which differ from each other in their goals, tasks, depth and time periods needed for achieving the stated goals.

Accordingly, one has to differentiate between MT as a scientific technical problem for designing a working MT system tested and used on a computer as a source of information, and MT (if that term should be used at all) for the various retrieval researches in which language studies are used for solving various processes in information treatment. In connection with this division, one has to evaluate various linguistic studies. MT as a universal scientific problem is a logical intersection of various sciences interested in aspects of language such as general linguistics, mathematical logic, semiotics, psychology, a series of cybernetic sciences, etc. Within this framework, the area of linguistic investigations keeps enlarging toward the more fundamental description of language disregarding its connection with the tasks appropriate for MT.

The studies and research aimed at designing MT systems as practical and working systems found themselves in quite a different situation. These efforts were not supported and as a result there is not a single working MT system in our country.

One of the basic causes of this situation is the conscientious and also unconscious mix-up and identification of MT as a scientific technical problem vs. MT as a universal scientific problem. As a result of the consequences of the "goals-tools" approach there was a disbalanced correlation between the levels of research serving the solution of MT as a scientific technical problem versus retrieval research disregarding any applied testing for MT.<sup>3</sup> The applied studies turned into retrieval studies, and having lost the connection with the original goal, naturally they were not able to secure or provide the solution for working MT systems.

Starting with '60s the development of MT in our country and the West went on different roads. One could completely agree here with the evaluation expressed by MT specialists.<sup>4</sup>

In the West, despite the discussion of "crisis" in MT, attempts to solve MT problems by "brute force," by using the great dictionaries and relatively simple algorithms were not stopped.<sup>5</sup> At the same time theoretical studies were also conducted concerning formalization of language structures in a mode of close connection with the computers. Thus, for example, the principles of syntactic analysis of sentences and generation of "microsentences" were programmed and experimentally tested.<sup>6</sup>

- 
3. G. Pospelov. Ob'ekt upravlenija-nauka., "NAUKA I ZHIZN," 1975, 11.
  4. V. N. Gerasimov, Yu. N. Marchuk, SOVREMENNOE SOSTAJONAIE MASHINNOGO PEREVODA, collection "MASHINNYJ PEREVOD I AVTOMATIZACSIJA INFORMACSI-ONNYKH PROCETSOV," M., 1975.
  5. J. M. Daniel, Translation by computer, "Electronics Weekly," 304, 1966, 7.
  6. B. T. Carmody, P. E. Jones, Automatic derivation of microsentences, "Communications of the ACM," 9, 6, 1966



the transformational model of syntax on the basis of Chomsky's concepts<sup>7</sup>, ways of graphic representations of the syntactic structures of language,<sup>8</sup> and other theoretical constructs in linguistics.

This provided the opportunity to evaluate the results of theory from the point of view of practical significance for solving MT tasks, to select the level of realistically necessary details and formalization of linguistic descriptions, to correctly modify the direction of further investigations.

One should not maintain the idea that the coupling of these studies with specific MT tasks, narrowing formal linguistic studies to their usefulness for MT, their feasibility for testing on the computer, hampered or suppressed the creative thinking of researchers and led to a blind alley, as is asserted by some purely theoretically oriented researchers. Experience showed the opposite. Combinability of the theory with the solution of practical problems and experimental testing of the theory led to the creation of the series of working systems, "unscientific" as their principles may be, with various degrees of automatization of the translation process. Thus, for example, there are in existence large automatic dictionaries, the use of which secures a higher quality and quicker human translation<sup>9</sup>, systems of translation for information data,<sup>10</sup> systems of MT producing "rough" translations of arbitrary texts, and with the additional editing--translations of high quality which is quicker and cheaper than by human hands.<sup>11</sup>

- 
7. B. J. Friedman, A computer system for transformational grammar, "Communications of the ACM," 12, 6, 1966.
  8. W. A. Woods, Transition network grammars for natural language analysis, "Communications of the ACM," 13, 5, 1970.
  9. H. J. Schock, Zusammenarbeit Mensch/Maschine beim Umgang mit elektronisch gespeicherten Informationen, "Zusammenarbeit Mensch/Maschine beim Umgang mit elektronisch gespeicherten Informationen," 1966.

Use of such a system provides not only practical help but also creates a realistic basis for further improvement of MT and development of retrieval investigations in the interest of perspective problems. These studies have demonstrated the role of linguistic investigations also for the development of electronic computing technology. In particular, they have made it possible to reformulate the new demands and conditions regarding the design of the computers of future generations including considerations of specific features characteristic of human handling of the information data.<sup>12</sup>

In our own country, the development of MT after 1961 took another road. A new trend in linguistic investigations was formed and rooted which considered MT as a general scientific problem, but retained the old applied title "automatic translation." Within the framework of this trend the scientific-technical problem of MT was pictured as one of the many specific problems the solution of which was possible after completing the whole complex of theoretical studies in linguistics. Retaining of the title "automatic translation" assisted this trend during its formative stage since it created an illusion that the research would continue for purposes of solving the scientific-technical task of MT, while in reality this new trend set as its purpose quite different tasks far removed from MT problems.

The fundamentals of this theoretical direction are most completely exemplified in the preface to the book "AUTOMATIC TRANSLATION 1949-1963."<sup>13</sup>

- 
12. D. G. Hays, Linguistics and the future of computations, "AFIPS Conference Proceedings," New York, 1973.
  13. I. A. Mel'chuk. Preface to the book: I. A. MEL'CHUK, R. D. RAVICH, AVTOMATICHESKIY PEREVOD 1949-63. A critical bibliographical manual, M. 1967. (Indications for pages are given in the text.)

In it, in particular, is said that the investigations in MT have as their aim formalizing linguistic descriptions and "perhaps, in a number of aspects the creation of a new science language"(p.7). "MT (automatic translation) is a specific job within a more general scientific goal--to teach the computers to learn human languages"(p.7). Thus, instead of solving a specific problem, a general problem with a global perspective is postulated. From this it follows that "the description of trends according to which the text is connected with the meaning is a central problem of linguistics--one that is theoretical and descriptive" (p.8). Within such wide phrasing of linguistic goals there is no natural border line between purely linguistic work and work concerning MT...Any sufficiently rigorous linguistic study or work that contains material adequately processed has a direct or at least indirect relevance for MT" (p. 9).

(It should be noted as an objective criterion for "strictness" and appropriateness of linguistic work for MT is the fact of its being included in the bibliographic list "AUTOMATIC ANALYSIS OF TEXT AND AUTOMATIC TRANSLATION" of RF (Referativnyj Zhurnal) "INFORMATIKA," whose editor is I. A. Mel'chuk himself.)<sup>14</sup>

Furthermore, all the questions of algorithmization of procedures, using the results of linguistic investigations for applied goals, and the experimental testing of linguistic algorithms are declared as not

---

14. I. A. Mel'chuk, OPUT TEORII LINGVISTICHESKIKH MODELEJ "SMYSL(--)  
TEXT," M., 1974 (Indications of pages are given in the text of this  
article.) English title: Experience in modeling the theory  
"MEANING(--)-TEXT".

needed since they only hinder the work of the researcher. This work, unnecessary for linguists to do, will be done by mathematicians who expect from the linguistics only abstract descriptions of language (p 9). In the final analysis, the formulation of the new trend proclaims "Can we have automatic translation without translation, without computers, without algorithms? Yes, such is the contemporary conceptualization of MT expressed in a somewhat paradoxical form." (p 9).

Assuming the above definition of "automatic translation," any declaration or mention of designing a working automatic MT model (i.e. creation of a system for automatic analysis and synthesis and, afterwards, also translation) (p 7) sounds like an empty pronouncement. At the same time, the continuous use of terms such as "automatic analysis," "automatic synthesis," "working model," etc. are understood in a very unusual way, only as an aggregate of rules, application of which are in no way connected by the researchers either with their algorithmatization or even more so with their testing on computers.

The central concept of this trend is the language model. In one of the last works in this trend the language model of the type (sense(--))text) is considered as the basic tool for investigation and description of language.<sup>15</sup> Essentially, the model "sense--text" is a description of procedures for linguistic structures on various levels, starting with the level of sense and proceeding to the level of its expression in a natural language, using the specially built formal languages, and the formal description of rules for transition from one level to another (the algorithmic

---

15. I. A. Mel'chuk, OPUT TEORII LINGVISTICHESKIKH MODELEJ "SMYSL(--))TEKST," M., 1974 (indications of pages are given in the article.)

part of the model). The applied value of these linguistic investigations is seen in the fact that after a more or less complete model "sense--text" will be built, it will be possible to store in the memory of the computer the system of formal description of all the levels of a natural language (in its totality!). This includes representation of the meaning of the linguistic units, and on the basis of the transition rules between levels one could design the corresponding algorithms and programs. Such a programmed mechanism will carry out the universal transformations between meaning and texts (modelling human speech behaviour). This is what is meant by the "language processor"--the universal transformer of information in a natural language. If two languages are described in this way, the problems of MT for this pair will also be solved" (pp 5-115).

However, the set goal--a global formal description of a natural language with any restrictions using a very loose concept of formalization and a lack of understanding as to what degree these descriptions could be used in the near future--was bound to lead the researchers to the creation of cumbersome designs difficult to inspect.

Applying these kinds of investigations to other languages and in depth analysis of specific problems makes the task even more complex and makes it doubtful that the initial set will ever be achieved by using these methods. Even if one abstains from considering the algorithmic aspects of models, i.e. description of rules of transition between levels of language description assumed for making the theoretical investigations easier (pp 5, 15, 27), their completion remains problematic. It leads to further expansion of the field of investigations, to even more detailed examination of linguistic descriptions, without taking into account their future use in the algorithmic part of the model, resulting in making the

perspective for adaptation of these investigations for practical uses less and less clear. One has also to take into consideration that even if the formal descriptions of language are completed, one would still need their application to the second part of the model "sense--text," i.e. the algorithmic one. This is connected with designing of the formal language of describing the rules for transition between levels and the algorithmic procedures themselves. In terms of complexity, this second part of the work is comparable with the first part, i.e. description of language levels.

The research using the model "sense--text" during the last 15 years resulted in the design of fragmentary and disconnected descriptions of language data, particular incomplete "models," a list of problems of principal importance to be solved in the future pertaining to the theory of language, a complete refusal to consider the algorithmic aspects of models, and recognition that the completion of these investigations is impossible within the foreseeable future. If, after this, one still hears assurances that the formalization of language with this method constitutes the only serious task of linguistics, it is appropriate to recall a remark of S. Lem in an analogical situation, "...they of course know that they will not success in a complete formalization of either the deductive or natural language, nevertheless they assume that the distance between 'an incomplete formalization' and 'a now yet incomplete one' constitutes a field that needs to be worked on for a long time with diligence."<sup>16</sup>

Under these circumstances one can not seriously see any immediate connection between research data and the specific scientific-technical

---

16. S. Lem, SUMMA TEKHNLOGII, M., 1968, p 231.

tasks at present and in the immediate future, getting results for solving the current problems of MT, informational retrieval, ASU (automatic control system), etc.<sup>17</sup>

These kind of assertions are not confirmed either by the actual state of investigations, or the results achieved at present time. Therefore, it is more proper to consider these investigations as purely theoretical, and assurances to solve in the near future the scientific-technical problems are unfounded, and the results so far achieved as an attempt "what degree of strictness is achievable at present in substantive linguistic descriptions."<sup>18</sup>

Thus, there developed in applied linguistics a theoretical trend with the goals and problems significantly removed from the scientific-technical tasks of MT. Essentially, it represents one of the theoretical trends of modern linguistics, and it is seen precisely as such in foreign literature.<sup>19</sup>

Evaluation of the significance of the ideas of this trend for the theory of linguistics is beyond the scope of this article. Here one should only find out its actual connection with MT problems since it is underscored with the retention of the title "Automatic Translation" and corresponding assertions of its representatives.

In terms of its applied significance this trend of investigations is unsatisfactory due to its immensity<sup>20</sup> and, accordingly, the impossibility

---

17. L. L. Iomdin, I. A. Mel'chuk, N. V. Pertsov, FRAGMENT MODELI RUSSKOGO POKERKHNOSTNOGO SINTAKSISA, I, PREDICATIVE SYNTAGMS, NTI, ser.2, 1975, 7, p.30.

18. Yu. A. Shreder, Predislovie k sborniku "Semiotika i informatika," 6, M., 1975, p 3.

19. "Trends in Soviet theoretical linguistics," D. Reidel Publ. Comp., Dordrecht--Holland/Boston, USA, 1972

of achieving the stated goal without narrowing the object of investigation and formalization. As far as the "strictness" of linguistic descriptions is concerned, there are some additional objections serious in nature beyond the already above mentioned "criterion": these objections refer to the description of the lexicon in the form of the explanatory-combinatory dictionary (tolkovo-kombinatornyj slovar')<sup>21</sup> as well as the extremely loose use of generally accepted terms and concepts.<sup>22</sup> One can hardly consider as an advantage of the trend under consideration. Its procedural techniques are aimed at linguistic universals and yet based on formal languages whose basis is constructed on logical constructs which are extralinguistic. There is no use of the "organization of the natural language," while one would expect that precisely this way might be the most productive.<sup>23</sup>

Statistical-approximate, algorithmic and other features of various language levels and text units turn out to be very useful both in constructing dictionaries, analysis and synthesis algorithms, as well as in description of the necessary data for them. Accordingly, it is quite fair to assert that the rational combinability of various approaches to use all the qualities of linguistic structures<sup>24</sup> is necessary.

- 
21. N. Z. Kotelova, ZNECHENIE SLOVA I EGO SOCHETAEMOST', L., 1975, P. N. Denisov, OCHERKI PO RUSSKOJ LEKSIKOLOGII I UCHEBNOJ LEKSIKOGRAFIJ, M., 1974.
  22. B. N. Golovin, LINGVISTICHESKIE TERMINY I LINGVISTICHESKIE IDEI, VJa, 1976, p 3.
  23. M. V. Arapov, V. B. Borshchev, Yu. A. Shreder, op. cit., p 15.
  24. N. D. Andreev, STATISTIKO-KOMBINATORNYE METODY V TEORETICHESKOM I PRIKLADNOM JAZYKOZNANII, L., 1976, pp 5-6.



Apparently, only by using such a complex approach will the perspective problem of universal transformation of information in a natural language be solved, i.e. the result will be a created "language processor" on a computer for multifaceted linguistic problems, including obtaining high quality MT.

The shortcomings of the trend under consideration does not imply that it does not deserve attention. The task of creating formal description is important for both theoretical and applied linguistics. It seems quite probable results that will have applied significance can be achieved by narrowing the goals, strictly delimiting the tasks and by restricting the subject of study. The point, however, is not whether the given trend is efficient. One should not agree with attempts to present it as the only correct path toward the solution of all theoretical and practical linguistic problems, especially since the connection of the problems, first of all, with MT is purely declarative.

The appearance of the immediate connection of this trend with MT is sustained with periodic reviews of history, state of affairs and perspectives of MT research both abroad and in our own country, as an example of which the preface to the collection of translations concerning MT could be cited.<sup>25</sup>

The concept of so called "three generations" in MT is presented in this overview: the first, the system of word-for-word translation; the second, the system of grammar; and the third, the system of semantic translation for MT. This concept represents an attempt to make an analogy between the development of computers and the development of MT systems. Although each of the computer generations differs from the

---

25. O. S. Kulagina, I. A. Mel'chuk, AVTOMATICHESKIJ PEREVOD: KRATKAJA ISTORIJA, SOVREMENNOJA SOSTOJANIE, VOZMOZHNYE PERSPEKTIVY, collection. "AVTOMATICHESKIJ PEREVOD," M., 1971.

preceding one by a different basis (tubes for the first, transistors for the second, and integrated solid systems for the third generation), architecture and technical characteristics, there are no generation in designing MT systems which is documented by their design abroad. The existing MT systems do not show the differences prescribed by this system and can be conditionally ascribed to one or null generation (the automatic dictionary of the Bundeswehr). The concept of the three generations does represent the real differences in the existing systems, rather than the evolution of views on the part of MT specialists on how MT could have been designed.<sup>26</sup>

Thus, it is considered that one of the principal differences of the third generation is the independence of the description of language material from the "mechanism," i.e. algorithms of translation. However, this principle is not generally accepted and what is even more important does not have any support from the practical work in MT. The researchers themselves who introduce this classification<sup>27</sup> consider it only as a way to improve the algorithm but not a goal in itself.

They emphasize the specificity of the translation process, requiring operations of linguistic units from various levels and consideration of interconnections between them which ipso facto does not make it possible to carry out this principle. The other specialists do not generally

---

26. See: V. N. Gerasimov, Yu. N. Marchuk, op. cited.

27. S. M. Lamb, STRATIFIKATSIIONNAJA LINGVISTIKA KAK OSNOVA MASHINNOGO PEREVODA, NTI, 1964, 10; V. INGVE, ZNACHENIE ISSLEDOVANIJ V OBLASTI MASHINNOGO PEREVODA, NTI, 1965, 7.

recognize this principle as sound since it leads to extensive complexity of algorithms and can not be justified by any advantages.<sup>28</sup>

One should note that the division of the language of transcription from the "mechanism" is a very difficult task even under the circumstances of designing a translator for converting the algorithmic languages into codes for computer commands,<sup>29</sup> though the algorithmic languages are relatively simple in comparison with the natural languages.

The other differences of the three generation classification turn out to be also non existent, such as the complete independence of the analysis from the synthesis, delimiting the levels of algorithms (for morphological, syntactic, and semantic analysis), differentiation of the description of linguistic data depending on the levels, the multiple choice routines, inclusion into the dictionary of information of encyclopedic type, etc. (p 20). In reality, as the experimental designing of a MT system shows, the carrying out of these principles in practice is attended with great difficulties.

This concept and, in particular, the principle of differentiating the linguistic description ("grammar") from the algorithms ("mechanism") served as the technical basis for the thesis advanced by the trend of "automatic translation," that the linguist should engage only in the description of the linguistic data, without getting interested in how his results could be used in practice. This in fact made it easier to turn the applied MT investigations into a part of theoretical linguistics. The consequences of this "unloading" of "applied" linguistics from algorithmic (applied)

---

28. P. Garvin, Algorithm of syntactic analysis "FULCRUM" (for the Russian Language), collection "Avtomaticheskij Perevod"; H. H. Josselson, Machine Translation in review, "Computer and automation," 8, 1968.

29. P. Intersol, Sintaksicheski orientirovannij transljator, M., 1969.

aspects of investigations immediately showed up. The literature on information retrieval and MT began to be filled with publications of theories of universal transformations, descriptions of "working" (only in the imagination of the their authors) language models and other "applied" (by name only) works. The leading specialist in the field of automation of information processes, M. Taube,<sup>30</sup> expressed 15 years ago some critical remarks aimed at the claim of "structural linguistics" to be able to solve MT problems. These remarks unfortunately are applicable also to these people. S. Lamb was especially caustic: The professional programmer knows what could be expected from computers and he knows the algorithmic limits of programs; yet he is surrounded by a "swarm" of specialists who do not help him at all in overcoming the difficulties, and reject them by their unsubstantiated statements."<sup>31</sup>

The results of research and the real attitude toward the task of designing a working MT system is best formulated by the authors of the "automatic translation" themselves in the above mentioned preface. In the beginning of the sixties, "It became clear that there must be carried out first of all the tedious and time consuming work on formalization of languages and only after that its results could be used for designing the MT systems."(p 12).

1971 (the results). "Studies of automatic translation became the source for a great number of varied and interesting works of a purely linguistic nature" (p 12). "...the base is created for a transition to a higher and

---

30. M. Taube, Vychislitel'nye mashiny i zdravyy smysl, M., 1964, pp 59-73.

31.

new stage in investigations of MT" (p 13).

1975. (The results of 15 years are stated in the book, "EXPERIENCE OF THE THEORY OF LINGUISTIC MODELS "SENSE--TEXT": "However, the stated goal at present is not feasible, at least for the author of these lines. Not a single full model "sense--text" is known to him even in a purely experimental nature..." "As far as the theory is concerned, even here there are tremendous lacunae....It is impossible to fill all these gaps in a foreseeable future (p 5). If one considers that the algorithmic aspect is not considered at all (p 15, 27), it is possible to evaluate this situation as a scientific blind alley of the theoretical trend in AT.

With such a situation in the field of "automatic translation" even in its theoretical part, not to speak of its applied aspects, one gets strange impressions from statements of some of their representatives regarding the futility of starting a new experimental MT prior to finishing the above stated theoretical goals since such new experiments would boil down to a repetition of the experiments of 20 years ago.<sup>32</sup> It is difficult to say what is the basis of such an attitude: lack of understanding of a new situation or a desire to defend one's own "scientific territory", including MT, at any price, from intrusion by practitioners.

It could not be maintained that nobody noticed the one-sided development of this trend in the field of MT. There were also warnings on the dangers caused by isolation from practical experiments. Thus, in 1964, G. G. Belonogov and R. G. Piotrovsky came out with an evaluation of the situation.<sup>33</sup>

---

32. For example, refer to the statements of V. V. Ivanov and I. A. Mel'chuk at the All-Union Conference on theoretical linguistic problems in 1974 (section "Theoretical Aspects of Applied Linguistics").

33. G. G. Belonogov, R. G. Piotrovsky, G. MASHENOV, REZHEKOV, M. 1964, 1.

In 1965 a book by D. Zhukov was published in which the history of MT was published, where the conflict between his applied and the theoretical trends was debated in sharply polemic terms, and the problems of designing working MT systems were discussed.<sup>34</sup> It should be noted that in the 1975 edition of this book, under the title "MY PEREVODCHIKI" (We Translators), the publishing house "ZNANIE" removed from the book all the criticisms against structuralism and the theoretical trend in MT without the agreement of the author.

In 1970 R. G. Piotrovsky, analyzing the causes of MT "crisis" in our country, wrote: "At the end of the fifties there were working MT collectives in our country, aiming at the creation of commercial MT and they were getting promising individual results. These collectives, overcoming technical and linguistic difficulties, could have probably achieved in the middle of the sixties commercial translations on new Soviet computers. Unfortunately, the absence of the technical base, the organizational and psychological atmosphere in mathematical linguistics and the turn toward theoretical automatic translation, emerging in the beginning of the sixties, paralyzed this work....Our lag behind the leading European and American collectives could be evaluated at 8-10 years."<sup>35</sup>

Considering this trend of "machine" translation "without translation, without algorithms," with its representatives evaluating all the work in

- 
34. D. Zhukov, PEREVODCHIK, ISTORIK, POET? SLOVO, TEBE, MASHINA, M., 1965.
35. R. G. Piotrovskij, OTRASLEVOJ VEROJATNOSTNYJ MASHINNYJ PEREVOD, collection. "STATISTIKA TEKSTA," II--Avtomaticheskaja prerabotka teksta, Minsk, 1970, p 7.

MT only from the point of view of their correspondence to the theoretical concepts of the trend, and when the success of the design of a practical MT system was made to depend on the completion of their "theory," there were only very few organizations which dared to carry out the practical MT work against the "scientifically" guaranteed failure. The theoretical trend itself could not produce the practical results and a vicious circle developed: a theory for theory's sake. It is precisely this point that we consider harmful, i.e. identification and confusion of the theoretical and applied MT work which led to the reduction of the applied work in MT and its substitution by purely theoretical investigations in the field of language.

The existing theoretical trend in linguistics has no right to call itself "automatic translation." The retrieval research in the field of language universals belongs to the area of theoretical linguistics and in terms of general scientific planning impinges on cybernetics problems.

These investigations at present and in the future have no direct relevance for the scientific technical MT problem. The retention of the title "automatic translation" makes it possible to assume that it is used, on the one hand, for prestigious reasons and protection of their ideas from criticisms on the part of other language specialists, and on the other hand, for the creation of the appearance of MT development in the USSR. It is difficult to find another explanation.

What then is the actual state of linguistics MT work and its nearest perspectives? As the work of the committee showed, the basic difficulties of MT work are linguistic (description of the language and procedures), mathematical (designing of algorithms and programs) and technical (computers and external devices) and are at present not insurmountable. Despite the

lack of support from representatives of "theoretical" translation, there was carried out in the USSR some practical research work, though not in sufficient degree. There exists quite extensive experience in building a large MT dictionary, algorithms for synthesis and analysis, programming and experiments on computers. All this is a good basis for the creation of practical MT systems. The organizational difficulties are also to a certain degree removed due to the formation of the All-UNION TRANSLATION CENTER/VSESOJUZNYJ TSENTR PEREVODOV (VTSP), which is charged with the responsibility of coordinating all the work in this field and creating working MT systems within the next few years.

The International MT seminar which took place in Moscow in 1975 in autumn showed that linguists are beginning again to do more and more practical MT research.<sup>36</sup>

The general direction of the research aimed at the creation of practical MT systems should be based on the actual available experience in formalization of the description of language data and designing of algorithms and programs, and should not be based on untested hypotheses and theories. The investigations and designs should adhere to the following principles:

- 1) orientation of translation systems for a pair of languages, for individual sublanguages, and if necessary for individual sets of information documents;
- 2) study and description of specific sublanguages with the degree of detail which is actually necessary and could be used in algorithms;

---

36. "MEZH DUNARODNYJ SEMINAR PO MASHINNOMU PEREVODU," M., 25-27 nojabrja 1975. Tezisy dokladov i soobshchenij," M., 1975.

3) general use of statistical data concerning the distribution of linguistic units in texts taking into consideration their environment and the properties of the 1) Approved For Release 2008/03/03 : CIA-RDP83M00171R001800120015-9



- 4) agreement and subordination of the language data description to algorithms;
- 5) designing of complete systems of algorithms and programs on the basis of the modular principle with consideration (within certain limits) for the possibility of changes within the linguistic information and algorithms, too;
- 6) using the existing tools of automation and programming for designing of experimental and testing programs and contemporary computers, the series ES EVM;
- 7) gaining accumulation of experience, experimental testing of partial plausible hypotheses and improvements of MT systems.<sup>37</sup>

In agreement with these principles, it is obviously necessary to carry out also the applied linguistic investigations in the field of MT, starting first with easier problems and as these are solved going on to more difficult ones. This is the usual path for solving any scientific technical problems, unfortunately forgotten in the "linguistics of AT" which assumes a universal solution of the problem first, in its totality (recognizing at the same time the unattainability of this goal) and only after that should one proceed with solving particular tasks.

Having said that much does not mean the lack of significance of theory and corresponding investigations and should not be understood as curtailment of theoretical investigations for the sake of present needs. The theoretical research should go on both in applied and general linguistics differing in goals, tasks and the duration needed for their

---

37. The strategy and tactics for solving the design of practical MT systems is contained in the article by V. N. Gerasimov and Yu. N. Marchuk, op. cit.

completion. However, there is no doubt that even without their completion there should be carried out research for designing the "approximate" methods of linguistic solutions for the MT system on the basis of technical tools now available and the present level of knowledge. The existence of even the simplest MT systems and their use permits timely testing of the intermediate results of theoretical constructs in the language area and correction of their further development. This kind of method should secure continuity in improvement of MT systems on the basis of the experience thus gained, tested theoretical results and new technical tools.

The contemporary state of linguistic research with a further continuing lack of experiments, i.e. the working MT systems, hampers not only its development, but also progress in the linguistic theory itself. This is the reason for symbiosis between theory and experiment and the strengthening of applied research in order to achieve the working MT system and solution of other urgent problems, relevant for the automatization of information processes in the national economy.

#### Chronical Notes

VJa, #5, pp 171-173.

An international seminar on MT took place on November 25-27 in Moscow, organized by the All-Union Center for Translation (AUTC)/ Vsesojuznyj Tsentr Perevodov nauchno-technicheskij literatury i dokumentatsii under the State Council of Ministers GKSM of the USSR on science and technology and The Academy of Sciences together with the International Center of scientific and technical information.

In the opening words, the representative of the GKNO (State Committee on Science and Technology), N. V. Turtanov, and the director of VTSP (AUTC), V. N. GERASIMOV, noted that for the development of machine translation it

is necessary to have both the experiments on computers using large linguistic data sets, and theoretical research in the field of applied linguistics and general linguistics.

The papers were grouped according to four basic topics: 1) machine dictionaries, 2) problems of automatic analysis and synthesis of texts, 3) semantic analysis of text, 4) mathematical and programming service.

The machine dictionary occupies a special place in the modern concept of MT. The dictionaries of natural languages within the MT systems and information retrieval systems have a specific design which makes them different from usual dictionaries. The basic characteristics of machine dictionaries are determined by their functioning in processing the language data: the autonomous nature is different from the complementarity of usual dictionaries, operational feedback, and interdependence and interaction with the algorithm of the automatic processing. Study and design of machine dictionaries is in the hands of specialists of computational lexicography as part of computational linguistics. Selective strategy in MT assumes transfer of "deep" difficulties into the "surface" level, and here a special role is assigned to the machine dictionary. A. Ljudskanov described in his paper the principles of such selective strategy consisting of the glossary of morphemes (People's Republic of Bulgaria).

A bilingual automatic dictionary of a thesaurus type represents a main link in the research "from linguistics of the text toward machine translation" within the group "STATISTICS OF SPEECH"/STATISTIKA RECHI.

These and other important questions were discussed in papers by A. Ljudskanov, R. G. Piotrovskij, Yu. N. Marchuka.

Many papers and reports presented specific results of already built and used machine dictionaries. (V. V. Goncharenki et al., L. N. Beljaeva, Z. M. Shal'japina, M. I. Otkupshchikova, G. Ya. Gordin et al., E. M. Dobruskina and others, K. B. Bektaev and others.)

Machine dictionaries play a great role in machine analysis of specific aspects in word formation and inflections, for example, in analysis of compound German words (V. A. Vertel').

The compilation of glossaries is preceded by lexicographic research with the aid of computers, and as a result the dictionaries-concordances (KWIC) are built, also semantic frequency dictionaries of parallel texts and backward dictionaries are produced (I. I. Ubin, E. V. Vertel', E. L. Koz'mina).

On the section devoted to automatic analysis and synthesis of texts the largest number of papers and reports were presented. The following statistics are interesting: among 17 papers presented, eight were devoted to automatic syntactic analysis, the morphological analysis, solution of lexical polysemy was dealt with in three papers each of which discussed the rest of the problems (the generative model and syntactic synthesis). It is apparent that the interest in syntactic analysis still prevails. The papers read are as a whole characterized by absence of new all-embracing theories, rather the research is carried out on the concepts already known ("meaning-text," statistical theories of speech and language) or here are constructed algorithms for specific problems without any kind of global theories (for example, algorithm of translation of numerals from decimal base in natural language and vice versa, A. F. Osyka).

The paper of I. Starke (Shtarke) German Democratic Republic (GDR) presented a method of transforming the syntactic constructions from Russian into German for translation purposes, applicable for various pairs of languages, and also for intralanguage transformations for identification/isolation of synonymic constructions. Representation of syntactic structures in the given system makes it feasible to solve such complex problems as finding the antecedents of pronouns (anaphoric) (G. Klimonov, GDR). G. S. Tsejtin presented new models for analysis, using the preference of linguistic constructs. These types of models, and also the models with limited nonprojectivity were tested experimentally (G. S. Tsejtin, B. M. Lejkina).

The collective paper of scientists from Czechoslovakia described the basic components of the functional generative model. Semantic representations (formulae) are generated based on the syntax of dependences (E. Beneshova and others). An attempt was made to make a choice of grammatical analysis of text on theoretical grounds.

The general principles set up by Hjemlev toward linguistic analysis are acceptable for the MT as well as the principles of descriptive linguistics. However, the concrete text analysis is based on its own laws which are best described by distributional-statistical methods supported by probabilistic evaluations. The main principle of this method consists in the fact that the main postulates are not given a priori, rather they are arrived at as a result of attempts at formalization of language (A. A. Koverin).

Many papers were devoted to specific syntactic problems; for example, interpretation of comparative relation in the grammar of syntactic analysis

of Russian texts (V. Yu. Morova), calculus of syntactic structures for the algorithm of analysis of the continuous type (E. E. Lovtskij), analysis of prepositions (N. V. Anikina, T. G. Nikanorova, N. N. Leont'eva), rules for designing predicate relations (M. S. Pershikova and others), The automatic morphological analysis of the written Polish language was treated in an interesting paper by A. Lukashovicha (Polish People's Republic). M. P. Muravitskaja delivered a report on automatic morphemic analysis of verb forms.

It could be said that in the semantic analysis the experimental method took root. Thus, the concepts of the semantic connectedness of the text are verified by algorithms for segmentation of texts into paragraphs and connected fragments (T. N. Rylova, L. V. Orlova, R. A. Kovalevich and others, T. V. Dolgaleva, G. S. Osipov and others). Some presentations deals with the dictionaries containing semantic information, and also the semantics of specific words and word combinations in natural language (M. I. Otkupshchikova, G. M. Il'in and others, O. A. Shteronova).

The research in the field of formalization of semantics provides the output of immediate interest for the information analysis and retrieval (automatic indexing and annotation, creation of dictionaries thesauri). Due to this the paper of V. A. Moskov'ov and Yu. S. Martemjanov dealing with a generative model was very interesting; the papers of communicative organization of text and its reflection in the semantic structures (E. I. Korolev, A. M. Shal'japina) and Linguistic Justification of the System "Question" "Answer" (Konrad P., GDR). The quantitative evaluation of the quality of translation (N. A. Kuzemskaja, E. F. Skorokhod'ko).

In the section on mathematical and programming services papers were presented dealing with the programming techniques of interaction between the machine and man within the process of inter- and post-editing in the MT systems (B. D. Tikhomirov and others), complex approach toward mathematical and programming services for MT (D. M. Skitnesvkij), and design of general and specific programs (N. A. Krupko and others, N. G. Arsent'eva, R. S. Karetnikov, L. N. Beljaeva, S. A. Anan'evskij and others, N. A. Balandinav, S. Krisevich and others, L. F. Lukjanenkov and others). The problems of automatic recognition and synthesis of hieroglyphs were also considered (S. M. Shevenko). The results of various programs were demonstrated, representing linguistic algorithms.

In the sessions of this seminar 205 persons participated, approximately 80 theses of papers were sent to the organization committee, and 63 papers and reports were presented in sections and at plenary sessions. Scientists and collectives from five countries and 15 cities of the USSR participated.

This shows the interest in MT and the trend of practically all Soviet collectives and scientists studying MT and also the specialists from other countries--members of MSNTI (?) toward organizational unity in designing scientific and technical problems for commercially suitable MT systems.

Yu. Marchuk (Moskva)