Union Calendar No. 895
86th Congress, 2d Session - - - - - House Report No. 2021

RESEARCH ON MECHANICAL TRANSLATION

REPORT

OF THE

COMMITTEE ON SCIENCE AND ASTRONAUTICS
U.S. HOUSE OF REPRESENTATIVES

EIGHTY-SIXTH CONGRESS
SECOND SESSION

Serial d

June 28, 1960.—Committed to the Committee on the Whole House on the State of the Union and ordered to be printed

UNITED STATES
GOVERNMENT PRINTING OFFICE
49006 WASHINGTON : 1960

RESEARCH ON MECHANICAL TRANSLATION

I. INTRODUCTION

Now the whole earth had one language and few words. And as men migrated in the east, they found a plain in the land of Shinar and settled there. And they said to one another, "Come, let us make bricks and burn them thoroughly." And they had brick for stone and bitumen for mortar. Then they said, "Come, let us build ourselves a city, and a tower with its top in the heavens, and let us make a name for ourselves, lest we be scattered abroad upon the face of the whole earth." And the Lord came down to see the city and the tower, which the sons of men had built. And the Lord said, "Behold, they are one people, and they have all one language; and this is only the beginning of what they will do; and nothing they propose to do will now be impossible for them. Come, let us go down, and there confuse their language, that they may not understand one another"s speech." So the Lord scattered them abroad from there over the face of all the earth, and they left off building the city. Therefore, its name was called Babel, because there the Lord confused the language of all the earth; and from there the Lord scattered them abroad over the face of all the earth.—Genesis XI.

Communications between many areas on Earth is very difficult and at times is almost impossible. Hundreds of languages and many more dialects place barriers between people that are more formidable than oceans and mountains. By manmade machines we can sail across the waters and fly over the mountains, but we have not yet devised machines to circumvent the language barriers. The mere fact that we have modern methods of travel that can move men from the Western to the Eastern Hemisphere in a quarter of a day highlights the confusion of tongues mentioned in the biblical story of Babel, and emphasizes the problem of languages which is a most important issue in our fast-moving society.

Wars have brought many countries in closer contact as evidenced by this generation's travels to most parts of the world in support of national objectives. War, too, has provided us with steppingstones, or means to language translation, as evidenced by the first spark of an idea to machine translation by Warren Weaver in 1946. During that year. Weaver suggested to A. D. Booth that all languages might contain basic elements which could be detected by means of the techniques developed during World War II for the breaking of enemy codes. Booth took the position that any digital computing machine having the necessary storage capacity could make a dictionary translation.[1]

As sparks often kindle fires, so these thoughts on mechanical or machine translation—affectionately referred to as "MT" by those in the business—formed the basis of the international research and development program now in progress.

Within a year Booth ana D. H. V. Britten, at the Institute for Advanced Study, Princeton, worked out a detailed "code" for use on an automatic machine to add a set of grammatical notes to the bare dictionary stem renderings.

[1] "Machine Translation of Languages," edited by William N. Locke and A. Donald Booth.

In July 1949, Warren Weaver wrote his memorandum entitled "Translation," which resulted in the initiation of active research projects at the following universities:

University of Washington.
University of California at Los Angeles.
Massachusetts Institute of Technology.

Reifler, of the University of Washington, in his 1950 "Studies of Mechanical Translation, No. 1," suggested the concept of pre- and post-editing. In 1951. Oswald and Fletcher, of the University of California, published their "Proposals for the Mechanical Resolution of German Syntax Patterns." In that same year Yehoshua Bar-Hillel became the first full-time paid research worker in machine translation at MIT.

Bar-Hillel organized and presided over the first conference on machine translation in the spring of 1952. This conference was made possible by a grant from the Rockefeller Foundation. Although conclusions were not drawn at this conference, there was general agreement that word frequency and word translation studies as a function of individual languages and scientific fields can be undertaken now; operational analysis of syntax with a view to programing in terms of machine operations can be started immediately.[2]

In January 1954, the Georgetown IBM experiment succeeded for the first time in effecting machine translation from Russian into English on a limited basis. This experiment used a vocabulary of 250 words and the six rules of syntax formulated by Prof. Leon Dostert, of Georgetown University. Paul Garvin, also of Georgetown, devised a code used by Peter Sheridan, of IBM, to program the problem.

A second major event took place in 1954, when the first issue of a new publication, Mechanical Translation, appeared. This journal is edited and published at MIT by W. N. Locke and Dr. Victor Yngve.

As the value of machine translation became more evident, additional Government agencies began sponsoring research, more colleges and universities became interested, and industrial organizations formed teams to provide services to this growing field.

A new discipline in research has been born from the students and teachers of languages, from the designers of computing machines and from the electronics experts who build and operate these complicated instruments. In the dawn of this era of machine translation it is well to repeat the words of Prof. Leon Dostert, when he said in his "Brief History of Machine Translation Research" that—

> One of the difficulties which has handicapped progress in the field of machine translation has been the lack of communication on the one hand, and the assumption of somewhat rigid positions on the other. To help remedy the first difficulty, we hope that the publication and distribution of our seminar work papers will be one step toward greater communication among the different research groups in this country and in England.
>
> In respect to the second problem, it is somewhat more difficult to suggest remedies. It seems to me that since we are still barely past the threshold of our investigation, it would be both premature and unscientific to cling narrowly to a given hypothesis or theory as to the most efficient manner in which the problem can be resolved.

[2] Ibid.

Purpose of investigation

During the first session of the 86th Congress, on May 25, 26, 28, June 2 and 17, 1959, the Committee on Science and Astronautics of the U.S. House of Representatives held hearings on "Dissemination of Scientific Information." The problem of translation of foreign scientific documents was discussed during these hearings. Mechanical translation was mentioned as an aid to the translation problem. The committee then decided to investigate the research program in machine translation in greater detail during the second session of this Congress.

Lt. Col. Francis J. Dillon, Jr., staff consultant to the committee conducted a survey of the research effort in machine translation prior to the scheduling of hearings on the subject. As the majority of the research is conducted with Government funds, it was determined that the principal witnesses should be the Government agencies concerned, along with the principal investigators or contractors. The five Government agencies sponsoring research are the National Science Foundation, Central Intelligence Agency, Army, Navy, and Air Force. The National Bureau of Standards is conducting research in this field at the request of the Army and is funded by the Army.

Hearings were held for 4 days. Because of the limited time available it was not possible to call all interested agencies and investigators. The Navy was not called. In addition to those mentioned above the committee heard testimony from the Massachusetts Institute of Technology, Harvard University, Georgetown University, Baird-Atomic, Inc., and International Business Machines Corp.

National Science Foundation

By 1954 there was enough interest in mechanical translation research to gain recognition from the National Science Foundation in the form of a grant to the Massachusetts Institute of Technology. Now, according to testimony presented by the National Science Foundation, 11 groups in the United States are engaged in various aspects of mechanical translation research with support from the Federal Government.

On September 2, 1958, the Congress passed Public Law 85-864 entitled "The National Defense Education Act of 1958." Section 901 of the act provided for a Science Information Service and directed the National Science Foundation to—

> undertake programs to develop new or improved methods, including mechanized systems for making scientific information available.

Section 10 of Executive Order 10807, dated March 17,1959, entitled "Federal Council for Science and Technology," states in pertinent part that—

> The National Science Foundation shall provide leadership in the effective coordination of the scientific information activities of the Federal Government with a view to improving the availability and dissemination of scientific information. * * *

Central Intelligence Agency

The first Government agency to show interest in machine translation was the Central Intelligence Agency. In 1951 their scientists discussed the possibility of developing an automatic indexing and translating machine with Dr. James Perry, then with the MIT Center for International Studies, and now Director of the Center for Documentation and Communications Research, Western Reserve University. After some preliminary work, Dr. Perry and CIA representatives, in June 1952, attended a meeting at MIT of linguists, logicians, and mathematicians on the subject of machine translation. The principal result of that meeting, which was promoted by Dr. Bar-Hillel and supported by the Rockefeller Foundation, was the further stimulation of interest and the realization of possibility in the minds of the linguists present.

In the next 2 years or so, CIA reviewed various proposals, including those from MIT, the Battelle Memorial Institute, and Georgetown University. Some of these were considered jointly with elements of the Department of Defense.

The CIA took the position during that period that the development of a machine translation capability was highly desirable, and should be supported. They also recognized that such a program had implications which transcended the interests of CIA and those of the intelligence community. Therefore, they considered it preferable that an organization with broader responsibilities than their own should be prevailed upon to take the initiative in pushing a comprehensive MT program. The CIA then identified its immediate need as a usable product, i.e., one which might well be far short of a perfect translation, but nevertheless highly useful. In return for an early MT capability to produce a usable product, they were willing to leave the achievement of superior results to a longer range program.

This pragmatic approach was the aim and purpose of the CIA in 1954, and remains their aim and purpose today.

CIA then negotiated with the National Science Foundation which culminated in early 1956, in an exchange of correspondence between the two Directors. The National Science Foundation agreed to administer any part of a program of research in machine translation which is agreed by all concerned to be desirable.

U.S. Air Force

In 1954, the Air Force sponsored its first research effort in machine translation. This initial effort involved experiments with an automatic dictionary of Russian to English. A magnetic drum and assorted electronic circuitry were used to look up meanings of Russian words. This work was done at the Harvard Computation Laboratory by Prof. Anthony Oettinger.

Early in 1955, the Rome Air Development Center of the Air Research and Development Command recognized automatic language translation as an area of major importance to the Air Force. As a result of planning, a technical approach was formulated for a complete automatic language translation complex, in which three major areas of research were defined. These involved automatic input, techniques and machines for translating the foreign language into English, and an automatic output.

Work in the second area required the greatest effort in research and development, and in May 1956, the Air Force started work with Dr. Gilbert King for research and development on his invention of the photoscopic disk. The photoscopic disk is to serve as the heart of the planned translation complex. At the same time research was started with Prof. Erwin Reifler at the University of Washington to direct the development of a Russian-English dictionary. This dictionary has been stored on the photoscopic disk.

Today the Air Force has 10 research and development contracts, 6 with universities and 4 with U.S. industrial concerns. Two of the contracts are with European universities.

U.S. Army

A relative newcomer to the field of research on machine translation—since 1958—is the U.S. Army. Two major programs were selected by the Army Research Office: one with the University of Texas, in German-English and English-German translation under Professor Lehman; and the other with the National Bureau of Standards in Russian-English translation under Mrs. Ida Rhodes.

National Bureau of Standards

The National Bureau of Standards turned its attention to mechanical language translation by use of an electronic computer upon the request of the Army. Their project on mechanical translation has been conducted as a service for and under the support of the Army. Work began in 1958 with a small group consisting of Mrs. Ida Rhodes, project leader, and three assistants.

The National Science Foundation has helped the Bureau of Standards by detailing Mr. Richard See to work with Mrs. Rhodes part time.

U.S. Navy

The U.S. Navy began research on machine translation in 1958. The Navy is a sponsor of research and as such, does not perform the actual research. The Navy approach is to support projects which lead to the development of new theories or techniques which can be used as tools or rules to solve practical problems. They have a contract with Wayne State University to concentrate on techniques which will enable translation of mathematical text. They also have a contract with Hydel, Inc., for the development of a computer memory which will enable more efficient use of computers for automatic translation.

Since the very inception of the high speed digital computer, the Navy has been heavily involved in the support of computers, computer organization, programing techniques, computer technology, and auxiliary computer devices. From June 1953 to August 1954 the Navy contracted with International Telemeter Corp. for the development of the large photomemory which is the heart of the I.B.M. machine translation system. Also from January 1954, to July 1957, the Navy Bureau of Ordnance supported fundamental research in pattern recognition at Baird-Atomic, Inc., which led to the print reader program of that company. With regard to computer output devices, the Navy has supported work which led to the now widely used Stromberg-Carlson high speed printer capable of printing 5,000 lines per minute. From 1952 through 1956 the Navy supported a basic study of Russian linguistics at Wayne State University. This resulted in a very prominent book by Prof. Harry Josselson, "The Russian Word Count."

II. Why Mechanical Translation

The world is divided by language barriers into about 4,000 linguistic communities. Many of these communities are small and represent primitive or underdeveloped cultures. But well over 50 of these language communities are large and important enough to carry on extensive trade, communication, and cultural interchange with one another. All interchange between language communities must now funnel through individuals who are to some extent bilingual. The resulting bottlenecks serve to stifle such intercourse and to keep the language communities in comparative isolation. So testified Dr. Victor H. Yngve, of MIT.

There are many reasons for conducting research on machine translation; however, the two most important reasons may be listed as: National intelligence purposes; and for the automation of languages to facilitate the dissemination of foreign scientific and other types of information to those who need it.

Prof. W. P. Lehman, of the University of Texas, states the need for mechanical translation in the following excerpt of his prepared statement for the record:

> Translation by machine is the only practicable method of meeting the urgent problem of making scientific and technical information published in foreign languages available to American science and technology. The volume of publications in numerous areas is increasing rapidly, as any survey demonstrates. Moreover, with the spread of scientific interest to all areas of the world, publications of value are being produced in a greater variety of languages than before. We cannot assume, even today, that significant publications will appear only in French, German, or English. Russian publications are now almost exclusively in Russian; Chinese scientists publish in Chinese; and as the level of technology increases in countries with a strong nationalistic tradition, the United Arab Republic, Indonesia, and so on, we can expect important works published in the national languages of these countries. If one moves from the technological sphere to that of intelligence, even the seven languages alluded to above are inadequate to keep abreast with activities today. They will not help us in all of South America, in much of Africa, nor in important sections of Asia. No one can master all the important languages while specializing in one of the sciences. For rapid access to publications these will have to be provided for him in his native language.
>
> At least for the present there is no likelihood that the number of important languages will diminish. For nationalism is complicating the world's linguistic situation. India, for example, is abandoning English in favor of Hindi. New countries in Africa may follow India's pattern.
>
> Nor will an artificial language—e.g., Esperanto or even an international natural language—solve the problem. All artificial languages have been based on those of Europe, so that they are not readily mastered by non-Europeans. Even the most widely used has a pitifully small number of speakers.
>
> On the other hand, in the nationalistic world of today, languages associated with great powers will be suspected for potential cultural domination. It is unlikely that native speakers of English will adopt Russian aз their technical language. We can expect the Russians to maintain a similar position toward English.
>
> A further possibility of keeping abreast with current publication would be employment of a corps of human translators. This possibility is dubious because of expense and unmanageability. Translators would have to master technical fields as well as different languages in order to produce adequate translations. The number of translators needed to deal with Russian nuclear physics would probably not be many fewer than the number of nuclear physicists oar universities have trained.
>
> We are left then with the necessity of devising mechanical means of translation, because of the number of languages that exist, the wide number of technical areas, and the tremendous volume of publication.

The Government agencies involved in mechanical translation research have stated the need for such a system as follows:

National Science Foundation

The National Science Foundation lists as its specific objective the automation of languages, but also states that there are broader implications that this research on language may have for the future of information processing, particularly in the fields of automatic indexing and abstracting and in the mechanization of systems for the storage and retrieval of scientific information.

The National Science Foundation also supports worthwhile research not supported by other agencies. However, it supports only those projects that are considered to be of scientific merit and that usefully supplement or complement other work in progress.

The Director of the NSF, in concluding his prepared statement, said:

> It is our belief that the problem of mechanical translation, involving as it does some of the most subtle aspects of human communication, is at the same time of the greatest importance to the scientific community and of the greatest difficulty, requiring years of intensive research for its solution. We are convinced that the promising results obtained thus far warrant continuation of support of research on mechanical translation by the (National Science Foundation and other Federal agencies.

Central Intelligence Agency

The intelligence requirements for national defense are the motivating forces of the Central Intelligence Agency in their conduct of a program of research on machine translation. The CIA included the following two short excerpts of a letter from Mr. Allen W. Dulles, Director of Central Intelligence, to Dr. Alan T. Waterman, Director of the NSF:

> I should like to reaffirm the deep interest which we in the intelligence field have in the possibility of translation of Russian language materials, particularly in scientific fields, into English by machine. In addition, many of us feel that the degree of human understanding that could be accomplished if language barriers could be lowered without sacrificing linguistic integrity might well be a major step toward peace. * * *
>
> It is our opinion that much is to be gained by the early development of a machine capability for translation, The national security can be well served if we have available the scientific and technical literature of the U.S.S.R. in English for detailed analysis as early after publication as possible. I am assured by leaders in electronic research that technological problems yet unsolved need not stand in the way of the rapid development of a machine once the linguistic research has been started.

In Russian scientific and technological literature alone, the volume has greatly increased in recent years. The CIA estimates that the annual available output is now about 780 million words. This increase has been accomplished by increased efforts of the Government to translate the most useful part of the production. And the performance, 95 percent of which is by the Government or under Government contract, is impressive. About 53 million words of Russian scientific literature are now being translated annually (of which CIA accounts for over 9 million).

The CIA witness presented the following seven reasons for conducting a machine translation program:

1. The volume of publications will continue to increase, and at a rate in excess of our ability to procure competent translations.

2. The quality of translation work done through contract arrangements is not uniformly excellent. Whatever the level of accomplishment in machine translation at any given time, the output is uniform. In short, machine translation holds out the promise of a uniformly more accurate product.

3. Machine translation also promises greater speed. We now give priority to categories and languages of greatest interest. Nonpriority items are invariably slow in reaching the reader. Perhaps the translator with the particular skill in a language, or in a discipline, cannot immediately take on the task. In any case, he cannot translate, on an average, more than 2,600 words per day. The machine can hurl them out at rates of 3,000 to 50,000 words per hour, depending on the computer used. And these rates will increase. Even if postediting were required, the man-machine system would appreciably outproduce the human translator working alone.

4. With machine translation, more translations would be available. This increased availability of translations would itself generate new and more widespread demands for them. We now strive to pass over only marginal material, but cannot be sure that we are invariably successful.

5. Greater availability would result in a better informed corps of scientists in this country. This would result in superior evaluations of scientific and economic developments in the [Sino-Soviet] bloc than is now possible.

6. The development of a two-way machine translation capability would make possible low cost production of American publications for sale in underdeveloped countries where low cost bloc publications now have an almost clear field to the detriment of U.S. interests.

7. And finally, the research done and the techniques developed for accomplishing translation by machine would contribute materially to the solution of problems in the broader field of information storage and retrieval, and the emerging field of language data processing.

U.S. Air Force

The Air Force is active in the development of an automatic language translation capability primarily because of its interest in the translation output for intelligence and research purposes.

Thus, the Air Force mechanical translation program is designed to provide a capability for continued assessment of foreign scientific and technical literature, both for the scientific and intelligence community. At the present time the large volume of foreign language literature results in a 5 to 8 months lag which prevents timely utilization of the information. The Air Force has formally recognized this basic need.

An understanding of "Why" the Air Force has undertaken a program of automatic language translation research is implied in the basic need discussed previously, as well as by the nature of the translation requirements of these agencies. The Assistant Chief of Staff, Intelligence, requires extensive translation of reports and publications. The Aeronautical Chart and Information Center requires translation of source documents that are used for the production of charts and topographical material. The Aerospace Technical Intelligence Center requires detailed translations of scientific literature in specialized fields. The Air Research and Development Command must stay abreast of the scientific progress of all foreign nations in all fields and use this knowledge in its own research programs.

At this point, a summary of the preliminary survey made by the Planning Research Corp. for the International Business Machines Corp. is included to accentuate the need for a machine translation capability.

> Estimates of the accelerating publication of foreign literature and increasing accessions of those publications indicate that U.S. Government requirements for translations will become steadily more acute, and that present human translation capabilities will become increasingly inadequate in the face of the demand. The 1959 world production of books, for example, was shown to be well over a quarter of a million titles already, and that almost 100,000 of these are published in the Soviet bloc. By 1970, the U.S.S.R. alone will produce 100,000 books, while bloc production may well reach 160,000 titles or 12 billion words. The relationship between foreign-language publication, U.S. accessions, intelligence community requirements and current translation efforts for the U.S.S.R. is summarized in the table below. It should be noted that although only about 5 percent of U.S.S.R. publications are accessioned by the U.S. Government, by 1970 the present rate of procurement will add about 2.5 billion words from Soviet books and periodicals to U.S. collections each year.
>
> Under the constraints of current human translation capabilities and costs, U.S. Intelligence Agency requirements are kept to an extremely modest level. Requirements total only 0.6 of 1 percent of Russian publications and 13 percent of current accessions. Current translation efforts are able to fulfill even less of the demand: 15 percent of the intelligence requirements; 2 percent of current accessions; and only 0.09 of 1 percent of Soviet production. Even if a more moderate criterion—the need to translate only critical intelligence information—were adopted, current accessions from the Soviet bloc and Communist China alone would require the translation of about 250 million words per month. Such a requirement would necessitate an expansion of the pool of private and governmental translators to over 50 times its present size, and when the translation of intelligence materials from areas outside the Sino-Soviet bloc are added to the burden, the costs in manpower and funds become prohibitive.
>
> Finally, it should be noted that the potential for machine translation of foreign languages developed in this study takes into consideration only requirements arising from intelligence activities in the U.S. Government. Translation requirements associated with other governmental activities and nongovernmental demands would expand the potential far beyond the present estimate.

Comparative volumes of Russian publications

	Books and monographs (million words)	Periodicals (million words)	Total (million words)	Proportional relationship (percent)
(*a*) Production, 1958-59	5,175	22,500	27,675	
(*b*) Production, 1970	7,500	56,000	63,500	
(*c*) Accessions, 1958-59	923	363	1,286	C/A±5
(*d*) Accessions, 1970	1,689	818	2,507	D/B=4
(*e*) Requirements, 1960	62	108	170	E/A=0.6 E/C = 13 F/E=15 F/C=2
(*f*) Translation, 1960.	13	13	26	

U.S. Army

Intelligence requirements, coupled with the Army's interest in scientific problems led the Army to support the activities of the National Bureau of Standards and the University of Texas in mechanical translation research. The Army witness stated that—

the Army, as you know, sir, has many uses for, and has been interested in automatic-data processing in a variety of ways for a number of years. In fact, I am sure you know that the Army was responsible for the development of the first high-speed computer.

In dealing with the scientific problems and the intelligence problems, it is a natural consequence of our computer interest that we would also be interested in this other use of computer techniques.

U.S. Navy

The reasons for Navy support of machine translation and auxiliary projects are three in number.

First, good machine translation would be of great and immediate value to the Office of Naval Intelligence. Much of the information used by ONI arrives in one foreign language or another. Acceptable automatic translation riot only would increase manyfold the amount of raw data which could be ingested, but also could improve the accuracy and consistency of available English translations. Perhaps most important, competent analysts would be freed from the necessity of personally translating documents which they need quickly, thus leaving additional time available for the more abstract aspects of intelligence analysis. An added quasi-intelligence benefit to the Navy resulting from the availability of good mechanical translation equipment would be the ease of translating information from English to the languages of the various foreign personnel encountered by naval forces in various parts of the world.

Second, the Navy has a very great interest in the translation to English of foreign scientific and engineering literature. The Navy spends many millions of dollars annually in discovering and developing new devices and methods. Wide availability of pertinent foreign information could easily shorten the development periods required and reduce the money spent on work already accomplished elsewhere. American scientists would accordingly be considerably more aware of foreign projects and would have a broader base of scientific research upon which to draw in attacking specific Navy problems.

Third, machine translation is a most exciting application of high-speed computer technology. As has been mentioned, many of the problems involved in machine translation are common to a number of fields of information processing, so that progress in one field results in progress in the others, as well as giving additional insights into the solution of problems in other fields. Thus, for example, advances in machine translation will yield simultaneous improvements to document storage and retrieval, high-speed data processing, and automatic programing. All these subjects are of vital interest to the Navy and contributions to them will in many cases have immediate and widespread application.

III. Foreign Research on Mechanical Translation

Soviet Union

The Georgetown University-International Business Machines experiment in January 1954 received widespread notice in the press and other media of communication. Early in 1956, in the journal, Problems of Linguistics, Soviet experts reviewed the Georgetown-I.B.M. experiment—which in contrast to some of the reviews which appeared in this country-—was remarkable and commendable for its objectivity. They announced at that time that they, on the basis of the information gathered from the Georgetown-IBM experiment, had started research in the field of machine translation, which they claimed had brought them beyond the level of achievement demonstrated in the Georgetown-IBM test.

Scientists in the Soviet Union began research in this field in 1955, conducting experiments on a computer at the U.S.S.R. Academy of Sciences. Both the Institute of Precision Mechanics and Computing Technique and the Steklov Mathematics Institute of the Academy of Sciences began research in this field at that time. Later, the Institute-of Linguistics of the Academy, Leningrad University, and other institutions entered this research field. In May 1958, the First All-Union Conference on Machine Translation was held in Moscow.

In April 1959, a conference on mathematical linguistics was held in Leningrad which dealt largely with mechanical translation research. Soviet work in this field has been largely theoretical up to the present time, and very few experiments with computers have been mentioned in their literature. This is in contrast with the research in the United States, where experimentation with computers often has played a central role in the research process.

The CIA witness testified as follows:

> When MT is discussed there is invariably an expressed interest m what the Soviet Union is doing in this field. I will not dwell on this except to say that the Soviets have a program which considerably exceeds our own in scope and size, and that they are doing very good theoretical work, though restrictions on the availability of computer time has limited opportunities to apply theory to practice. Two papers, one by Professor Oettinger (Anthony G. Oettinger, "A Survey of Soviet Work on Automatic Translation," "Mechanical Translation," vol. 5 No. 3, December 1958, pp. 101-110), and one by Dr. Harper (K. E. Harper, "'Soviet Research in Machine Translation," Rand Corp. Monograph No. P-1896, Feb. 4, 1960, 17 pp.), provide valuable assessments of the Soviet effort. The Joint Publication Research Service series "Soviet Developments in Information Processing and Machine Translation," will also be of interest to the committee.

The Soviet effort is of more recent origin and in fact is going ahead much more rapidly. There are in excess of 80 institutions in the

Soviet Union that have programs in this field, including the following: Institutes of the Academy of Sciences of the U.S.S.R., and the Union Republic Academies of Science. They also have major programs, for example, at the Institute of Precise Mechanics (Moscow), the Electro-modelling Laboratory of the All-Union Institute of Scientific and Technical Information (Moscow), the Steklov Institute of Mathematics (Moscow), and the Experimental Laboratory for Machine Translation (Leningrad).

Professor Dostert, from Georgetown University, stated the U.S.S.R. effort was more advanced than the CIA witness indicated. He submitted for the record that the Russian effort in the field of linguistics and related sciences is far in excess of the U.S. effort. In the field of machine translation alone, evidence indicates that between 700 and 1,000 specialists in languages, linguistics, mathematics, compilation techniques, and engineering are at work in the Soviet Union. There are indications that the work on machine translation is focused on approximately 50 languages and that in respect to some of them, their system is operative. A basic advantage which helps Soviet activities is effective coordination of the diversified aspects of the total effort.

Professor Dostert received a paper from Andreyev of the U.S.S.R. Academy of Sciences, entitled "Basic Problems in Applied Linguistics", which made the following points:

1. *Increasing importance of linguistics.*—The increasing importance of language communication necessitates increasing attention both to theory as well **as** practical aspects of linguistics or the science of language. Although applied linguistics in the Soviet Union has made great strides, it is still lagging behind other sciences.

2. *Development of alphabets.*—Soviet linguists have succeeded in developing alphabets for the people of the U.S.S.R. who have no written language. The same effort is being made by linguists in southwest China. Africa has a long standing problem in this field.

3. *Language teaching.*—The importance of linguistics in language teaching methodology is increasingly recognized. Morphological and syntactic portions of algorithms developed for machine translation can be successfully used in language Instruction.

4. *Transcription and transliteration.*—A uniform transcription and transliteration system must be developed.

5. *Emerging scientific terminology.*—The increasing growth of scientific terminology through creation of new words for new concepts makes it necessary for a major effort to be made in lexicology, both in terms of creating new words and standardizing terminologies of diverse disciplines.

6. *Translation of scientific texts.*—Translation of scientific texts is subject to linguistic laws rather than being a problem of aesthetics.

7. *Shorthand*.- —Improved stenographic systems can be developed on the basis of new data derived from information theory.

8. *Speech defects.*—Linguists and psychologists can contribute to the improvement of speech defects.

9. *Orthoepy.*—The importance of developing orthoepy on the basis of the methodology of linguistics for greater facility of communication is essential.

10. *Communication channels.*—It is necessary to insure their effectiveness on the basis of rigorous linguistic analysis of the accuracy of the messages carried.

11. *Compression of speech.—G*reater economy in communication system can be achieved through this means.

12. *Compression of written speech.*—At present, lexical coding is, being developed to provide compression of telegraphic codes to about one-fourth their original length. The importance of written language compression is obvious for the field of machine translation and data processing.

13. *Transposition of codes.*—The use of computing machines should be increased for the transposition of codes and for extracting information in linguistic form.

14. *Scanning and listening devices.*—Research should be accelerated for the development of electronic scanners, both for visual and auditory perception.

15. *Speech synthesis.*—Greater emphasis should be given to speech synthesis in view of its importance for oral machine translation.

16. *Machine translation.*—Even though the present efforts are increasingly important, machine translation must make greater strides not only for its practical value, but for the information it will yield on the overall theory of language.

Great Britain

British scientists have long been interested in mechanical translation research, but full-scale research began only in 1955 when a grant for this purpose was made to Birbeck College, University of London, by the Nuffield Foundation. In March 1957, the National Science Foundation and the Rome Air Development Center of the U.S. Air Force began joint support of the Cambridge Language Research Unit in its mechanical translation research. In the spring of 1959, research in this field was begun by the National Physical Laboratory, at Teddington, England, an organization that is roughly comparable to our National Bureau of Standards.

Italy

A group at the University of Milan in Italy has been studying this problem for a number of years, and since February 1959, has been supported by a contract with the Rome Air Development Center.

France

In France, researchers interested in mechanical translation have recently formed an association for the study of problems in automatic translation and applied linguistics. In December 1959, a study center for automatic translation was established by the French National Center for Scientific Research. A member of the U.S. Air Force serves as the U.S. representative on the mutual weapons development team, for exchange of technical data on automatic language translation research with France.

Japan

In Japan, the Electrotechnical Laboratory of the Japanese Government, Tokyo, has been conducting research in mechanical translation.

China

Finally, according to a recent article in a Soviet journal, research in this field has been carried out in Communist China since *1958.*

IV. THE RESEARCH PROBLEM

General

Dr. Burton *W.* Adkinson, Head, Office of Science Information Service, National Science Foundation, introduced his prepared statement to the subcommittee with a discussion of the different approaches and objectives in the field of machine translation research. He stated:

Although the work in this field is often described simply as mechanical translation research, there are actually wide differences as Dr. Waterman said, in methodology, subject matter, and objectives among the various research groups. The research of some groups, for example, centers around computer experiments. On the other hand, some groups have had no recourse to computers at all, or have made only incidental use of such equipment. This difference in the use of equipment often reflects wide differences in the nature of the research. Even among those groups that use computers there is an important difference in approach or methodology. Some groups use a computer merely to verify that the procedures they have worked out function properly. Others prefer to use computers as a means of carrying out experiments with natural languages in order to learn more about language itself. These two uses of computers may be kept quite distinct in some groups and blended in others, but in any case are indicative of important differences of approach.

Another difference among groups is in the languages under study. Most of the groups in this country are studying translation from Russian to English because the practical need here is great. The large amount of Russian-to-English translation now being done by human translators is indicative of this need. Some groups, however, are also studying the problems involved in translation from other languages, such as French and German.

Of course, in other countries work is often concentrated on translation into the native language of the researcher, with the corollary that exchange of results on an international basis is often of interest largely from a theoretical or abstract point of view, since the actual detail with respect to any one language may be of little direct interest to a researcher studying another language.

At the present time it is impossible to say to what extent methods for translating the literature of, say, physics will be applicable to translation of articles in some other field, say biology. It is clear that there are at least some differences in vocabulary which may make it necessary to study each discipline in which the researcher is interested if machine translation is ever to be achieved on anything comparable to the human level. At present in the United States, the groups studying the problem of Russian-to-English translation are concentrating on such disciplines as physics, electronics, mathematics, chemistry, and biochemistry.

Groups differ greatly in their objectives. Some are aiming at a crude and yet useful product in the most immediate future. Others are interested only in high-quality translation by machine requiring no polishing or editing for its use. Regardless of the immediate objectives, however, the ultimate goal of mechanical translation research is complete automation of the process of translation, and we, in the Foundation, regard the terms “mechanical translation,” “machine translation,” and “automatic translation” as referring to this goal of completely automated translation. The work thus far indicates that there is a strong possibility that this goal will one day be attained. Much more research is needed, however, to determine whether or not this ultimate goal is indeed possible, and if possible, economically feasible.

It is possible that the crude machine output that can be produced at this time, consisting mostly of ungrammatical sequences of words, might be found useful by some organizations as indicative of the content of processed material. We do not believe, however, that such unedited output will be useful to research scientists.

There remains the additional possibility that machines may somehow be used to aid human translators in their work, so that the resulting man-machine complex will be able to translate either faster than a human expert, or more economically, or both. We understand that consideration is being given to the possibility of utilizing some of the intermediate research results and procedures to produce machine output that would then be converted by humans into usable translations. We have as yet, however, seen no convincing evidence that such partial automation of the translation process at this time would be an improvement over existing human translation. Before any sound conclusions can be reached concerning the usefulness of partially mechanized translation, there is need for study and objective evaluation of—

(1) The quality of the machine output achieved by mechanized procedures developed by several of the research groups, and

(2) The amount of human effort required to convert the machine output into usable translations.

Some of the research groups and sponsoring agencies have a still broader objective and are interested in mechanical translation as one aspect of a much larger problem, that of processing natural language by machine for a wide variety of purposes, including automatic abstracting and automatic indexing for storage and retrieval systems.

Dr. Victor Yngve, of MIT, stated that---

> work in mechanical translation can be separated into three parts: Science, technology, and production. Under science, we have research directed toward the discovery of the basic facts and knowledge of languages and translations that will form a firm foundation for erecting a technology. Under technology we can include research leading to the development of the dictionaries and machines that our science tells us how to build. Under production we would, of course, contemplate actual use of the technology for the production of useful translations.

Dr. Anthony G. Oettinger, of Harvard University, continued by stating that—

> like all forms of automatic information processing, the process of automatic translation may be divided into three phases: Input, logical processing, and output * * *.

The problems in a research program of this type are many. As the program is comparatively new, research is progressing along the three avenues outlined by Dr. Yngve and we see different organizations pursuing courses which can be classified as science, technology, and production. We also see each of these classifications being applied to the three phases of machine translation research as listed by Dr. Oettinger: Input, logical processing, and output.

Most of the scientific research is centered on the logical processing phase of this problem, as the input and output problems are mainly or a technological nature. For experimental purposes in the logical processing field, there is no urgent requirement for high-speed input and output. However, for a semioperational or fully operational capability, high-speed input and output are not only desired, but required.

The Air Force is conducting a large program in mechanical translation, which is equipment oriented, with the objective of attaining a complete automatic language translation complex. The Air Force program includes development of input, processing, and output machines. The Georgetown University is ready to start an elaborate program of translation of Russian documents by their machine process, but will rely on a manual key punching of cards for their input. Other agencies are devoting the majority of their research effort to the solution of the logical processing phase.

First, regarding input, the need is for a print reader capable of recognizing all types of fonts used in Russian literature, accommodating charts and pictures, and doing this job at as high a speed as. the processing machine operates. This subcommittee heard testimony from only one contractor—Baird Atomic, Inc.—on this subject. The Baird Atomic machine will have the speed, but is not yet ready to handle charts and photographs. Most of the witnesses testified that they believe a print-reading capability will be available when the logical processing phase is ready.

Second, regarding the output, there are many commercial products capable of providing a high-speed output for printed text. The task is to combine this high-speed printing output with the original equations, charts, graphs, and pictures, so as to maintain the original format.

The third, and major problem, is the logical processing. Here there are many differences of opinion, which is normal, if research is to progress. It would be a stagnant field if all agreed to one course and pursued that course.

Dr. Oettinger, of Harvard University, in discussing this third point—which includes syntactic analysis, said that—

> syntactic analysis still presents many challenging experimental and theoretical problems.

Proceeding with his prepared statement, he gave an example of a Russian word and the syntactic and semantic problems associated with resolving its proper meaning. Syntactic analysis can resolve some of the problems associated with a word, but figuring out meaning must be left to semantic analysis. Dr. Oettinger stated—

> semantic analysis deals with the elusive concept of meaning, and little that is both worthwhile and correct can be said about it at this time. Beautiful, smooth-looking translations can be produced here and now by various techniques all either unsafe or question begging. For instance, if precisely one English correspondent is stored in the dictionary for each Russian word, the problem of choosing among several correspondents obviously disappears; or, alternatively, if correspondents are labeled by technical field and the one most likely to occur in a given text in a given field is chosen invariably, then again the problem disappears. In either case, errors as yet unpredictable in number or in effect will occur, but the reader is led into thinking that because the result produced by the machine is smooth English, it must be right. The dangers of such a situation are obvious and similar to those inherent in the employment of incompetent human translators. Another possible technique is to interpose between the machine and the consumer a corps of bilingual technically competent editors, responsible for checking the work of the machine. Such a technique obviously begs the question, and such a staff might well be more profitably employed translating in the old-fashioned way.

Dr. Edward W. Cannon, of the National Bureau of Standards, stated :

> Considering the formidable difficulties which face the human translator, we must exercise extreme care in attempting to use for this task a manmade device which possesses neither the senses, nor the brain, nor the lifetime of experience bestowed upon man. For example, the conscientious human translator frequently augments the information stored in his mind by the use of grammars, scientific texts, dictionaries, and special glossaries. This implies that for automatic machine translation the vast totality of information, whether inherent to the human mind or stored in the accumulated literature, must somehow be fed into the electronic processor. Not only do we lack, at the present time, adequate devices for such a task, but the expenditure of time, labor, and money to store such colossal amounts of data will be prohibitive for many decades to come. We must therefore lower our sights and make a judicious choice of the data to be fed into a translation machine. But this choice depends both on the mechanical translation scheme which is being evolved and on the type of equipment which will eventually be used. The task of the NBS Mechanical Translation Group thus assumes a double aspect—to strive to achieve a workable mechanical translation scheme and, simultaneously, to keep abreast of and be prepared to utilize promptly engineering developments.

Postediting

Discussion of the research problem must include a few remarks on one of the major controversial issues surrounding the problem of mechanical translation: postediting.

Brig. Gen. William S. Ely, the Army witness, stated that their meaning of machine translation is—

> 90 to 95 percent accurate transfer of the intended idea, counted by whole sentences, with no preediting or postediting. If the product serves its purpose in this manner, it is truly machine translation. If it requires postediting of any sort, it should be called machine aided translation.

Discussing this same subject, Dr. Cannon, of the National Bureau of Standards, said:

I think that the requirements for effective utilization of such a procedure, the preservation of meaning from the source to the target language, en what we call posteditors, are more stringent—at least as stringent—than those on the human translator. I believe that it would be difficult to do the job, and I doubt that the output would be as fast—well, I think the output, the rate of translation, if the posteditors conscientiously should strive for accuracy, would be no greater than that of the human translator.

Prof. L. E. Dostert took exception with the witnesses who preceded him by stating:

In my years of experience in organizing and directing translation staffs, I have never known or heard of one, whether in the U.S. Government or in international agencies, that did not provide for revision. Yet, we have been told that revision in the case of machine translation renders the machine output trivial.

Mr. Robert F. Samson, of the Air Force's Rome Air Development Center, said:

I would like to say, first off, I agree with Professor Dostert. Human translation or machine translation will always require some postediting, but I feel this will be done by the reader. Because there exists no objective yardstick today for measuring the quality of translation output, one cannot say how much postediting is actually required in any particular case. I would say, however, postediting is done to improve the flow of words for easy reading. I feel this is somewhat of a luxury at the present time.

When asked about the accuracy required in translations, Dr. Gilbert W. King, of I.B.M., stated :

Speaking as a scientist, I, myself, could read output perhaps that ought to be somewhat better than we are getting now, but I do not feel we have to have it 100 percent accurate.

Profs. Harry H. Josselson and Arvid W. Jacobson, of Wayne State University, in their prepared statement for the record, stated :

We believe that postediting in actual production translation work will be necessary, but, as more experience is gained and procedures are refined, the amount of postediting will diminish.

This briefly gives the story of postediting of machine translation output.

Special purpose machine

The design of a special purpose computer for language translation purposes is an item of conflicting opinion. In general, most witnesses stated that the present general purpose computers are suitable for automatic translation research while the techniques for the ultimate system are developed. Most felt that the general purpose computer would not be suitable for the large mass of language material, mainly because of economic reasons.

The I.B.M. witness, whose company is the manufacturer of large computers, while discussing general purpose computers, stated:

We find that it has only limited use, because machines of this type are not really suitable as they are. They are really computers, and there isn't anything to compute in the language problem.

Others expressed the opinion that the standard computer would be best, as they are located around the country and thus will allow language translation at any place that has the program.

The cost of using a standard computer would be higher, as all the features of the standard computer not required for language translation would be purchased although not needed or used. When the program for the machine is perfected, the translation requirements will be high and computers will have a full time use in the national translation program.

V. APPROACHES

As in most problems of research we find that various investigators take different approaches to the problem. Machine translation research is no different. In classifying approaches to this research problem, the various types have been referred to as statistical, systematic, empirical, objective, and subjective. Definitions of these approaches may overlap. In his addition for the record of the hearings, Dr. Cannon, of the National Bureau of Standards, made the following remarks:

> When the idea of automatic translation emerged, at first only word-for-word translation was considered. The inadequacies of this plan were obvious. They manifested themselves principally in the facts that some source words had more than one target word associated with them, and that the word order in the target language often has to be different from that in the source language. Speaking of multiple-target words and of word order, however, is not the most useful approach to classification of problems. The change in word order can be understood by analyzing the grammar of a sentence. Multiple-target words can be either different grammatical forms belonging to the same stem or words having entirely different meanings. Thus, in the main, the problems encountered in translation are classified into syntactic and semantic problems.
>
> The three approaches to the semantic problem which we have just outlined might be named the statistical, systematic, and empirical approach. The same three methods of approach can be distinguished in dealing with the syntactic problem.
>
> Here, the statistical approach consists in searching through large amounts of source texts and enumerating the frequency of certain word sequences. For instance, how often does an adjective precede a noun, and how often does it follow it? The systematic approach attempts to set up a system of rules—in other words, a machine program—which analyzes the syntactic structure of each source sentence. That is to say, it identifies subject, predicate, direct object, etc., of each sentence or clause. This frequently has to be preceded by a grammatical analysis of each word, just as conventional grammar is divided into morphology and syntax, the former dealing with the inflectional forms of each word, the latter with the function of each word in the sentence. Finally, the empirical approach starts by selecting a few very simple rules for translation, tries them on a body of text and notices where they fail, corrects the rules or introduces new ones to cope with the observed failures, tries the revised rules on a larger body of text, and so forth.
>
> From a slightly different viewpoint we may distinguish between the use of conventional grammar and the design of new systems of grammar (or linguistic structures, as they are called), which are intended to be better suited to mechanical analysis than is conventional grammar.
>
> There are other differences among the various groups working on machine translation. For instance, in designing the dictionary or glossary, some propose to list in the glossary every inflectional form of every source word, while others propose to list only the stem, or equivalently some canonical form such as the infinitive of a verb and the nominative singular of a noun. Russian nouns have a dozen inflectional forms, adjectives and verbs many more. Thus the size of the required glossary is greatly affected by this decision.
>
> Other differences are found in limiting the scope of a machine translation project. Some groups are satisfied to translate into a kind of pidgin English. Some are resigned to leaving certain semantic ambiguities unresolved and printing out multiple meanings. Some are willing to admit failure in a small percentage of all cases. Some will even admit undetected errors in the translation, a point of view which others consider dangerous. Some propose to use a man-machine partnership rather than letting the machine do the entire job. In these cases the machine prepares certain aids to translation—at best a kind of preliminary draft, and these are used by a "posteditor" in producing a polished

translation. Preeditors, are less frequently contemplated, but a certain amount of preediting may be combined with manual key punching of the text

On one extreme is the approach taken by some that only 100-percent perfection is acceptable. Dr. Yngve stated that—

any quoted percent accuracy means very little, however, because of the difficulty of assessing less-than-perfect output. The trouble is that, even if errors could be counted, it is difficult to determine the relative amount of loss caused by different kinds of errors. Some errors are not serious, others are very serious. But, if I were pressed for a figure, I would say that realistically we can't reach an accuracy of 50 percent at present. But even if we could achieve 95-percent accuracy, what would it mean? Would it mean that we would miss the 5 percent of important new material and get the 95 percent of already known material? There is some indication that this would be the case.

The approach taken by Prof. Leon E. Dostert, director of the machine translation research program at Georgetown University, is empirical. He stated that after participating in the machine translating meeting at MIT in June 1952, he came away convinced that there was enough promise to warrant an empirical approach to the problem. He stressed the word "empirical" because the vastness and complexity of a general theoretical solution became quite apparent in the course of the discussions; and he felt that a pragmatic cumulative approach might prove more fruitful. Professor Dostert compared his cumulative, empirical-pragmatic experience to Lindbergh's flight of the Atlantic in 1927, in a monoplane, alone, to today's 90 or 100 people in a jet plane—that did not invalidate the importance of the 1927 flight.

Another approach is that taken by Dr. Gilbert King, director of experimental systems research of the International Business Machines Corp. Dr. King stated:

It has been our aim in automatic translation to consider all aspects of the problem and to make use of the skills and backgrounds of men who have pioneered before in the application of machines to this kind of nonnumerical material.

At present, Dr. King's group has developed a method to translate phrase by phrase and is working to extend the range of the clues from the phrases to the whole sentence.

Thus, the problem centers on the method of machine translation and extends from word look-up in an automatic dictionary, to phrase look-up or composition, to the full sentence look-up and rearrangement. The requirement for preediting and postediting is frequently discussed. Syntactical analysis of the sentence appears to be one of the agreed upon goals of mechanical translation.

Best known for syntactical analysis or predictive analysis approach is the group at the National Bureau of Standards under the leadership of Mrs. Ida Rhodes, and sponsored by the U.S. Army. Mrs. Rhodes' approach can be classified as subjective or systematic. Dr. Oettinger's group at Harvard University is following a similar approach to Mrs. Rhodes' group.

However, the best results expected from predictive analysis call for "Pidgin English," with correct meaning. The next major problem to be tackled is that of semantics.

Still another approach, described by General Ely, of the Army, is that of Dr. W. P. Lehman's project at the University of Texas. One might call this the objective approach. Examine the parallel text in two languages and derive rules for translation by means of the computer. This is in effect the systematic approach, referred to by Dr. Cannon, but it has the advantage of being equally applicable in either direction for a pair of languages.

VI. THE RESEARCH PROGRAMS

Funding

In Government the method most commonly used for measurement of effort is the dollars expended. A summary of the effort of the five agencies concerned with funding research in mechanical translation is as follows:

	All prior years	Present year	Next year
Army	$109,000	$170,000	$225,000
Navy	50,000	50,000	70,000
Air Force	3,400,000	1,400,000	1,500,000
CIA	315,000	177,000	[1] 922,000
NSF	1,063,300	325,000	490.000
Total	4,937,300	2,122,000	3,207,000

[1] Proposed.

National Science Foundation

The National Science Foundation does not conduct research, it sponsors a research program through grants to various organizations. The NSF grants for research on mechanical translation for fiscal year 1955 through the third quarter of fiscal year 1960, are as follows:

Grantee institution	Date of grant	Duration of grant	Amount of grant	Amount transferred from other agencies
Massachusetts Institute of Technology	October 1954	1 year	$18,700	
	October 1955	do.	24,800	
	October 1956	do	35,200	
	September 1957	do	41,400	
	November 1958	do	90,600	
	October 1959	do	126,000	
Total			336,700	
Georgetowm University.	June 1956	1 year	100,000	[1] '$65,000
	June 1957	do	125,000	[1] 90,000
	June 1958	do	186,000	[1] 150,000
Total			411,000	305,000
Cambridge Language Research Unit	March 1957	l year	27,100	[2] 20,000
	December 1957	do	33,000	[2] 20,000
	April 1959	do	35,650	[2] 20,000
Total			95,750	60,000
Harvard University	January 1958	6 months	29,150	[2] 15,000
	June 1958	4months.	26,200	
	September 1958	1 year	220,000	[2] 70,000
	December 1959	do	200,000	[2] 100,000
Total			475,350	185,000
University of California	September 1958	1 year	40,500	
	June 1959	do	57,600	
Total			98,100	
Grand total			1,051,900	550,000

[1] From CIA. [2] From RADC.

NSF grant for research related to mechanical translation

Grantee institution	Date of grant	Duration of grant	Amount of grant
University of Pennsylvania (Syntactic analysis of English for information retrieval).	October 1956	3months	$1,950
	February 1957	1 year	24,300
	February 1958	16 months	42,300
	October 1958	6 months	31,450
	June 1959	2 years	321,800
Total			421,800

The Foundation does not give instructions to the research groups working under Foundation grants. The Foundation believes that the research can best be planned by the research scientists and technical specialists in each field of research. The NSF submitted for the record of the hearings a summary of plan of research and results achieved by each of the organizations that has received grants for mechanical translation research:

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Summary of plan of research

The objectives of the work on mechanical translation at MIT have remained essentially unchanged since initiation of the project in 1954. The primary objective of this basic research program is to find out how languages can be translated by machine. Secondary objectives are concerned with evaluating the fidelity which can be achieved with different approaches, the usefulness of the resulting translations for various purposes, and their respective costs. A further objective is to add to the general knowledge of noncomputational use of digital computing machinery and to a basic understanding of human communication.

Results achieved

In light of these objectives considerable progress has been made. After consideration of the fidelity that could be achieved by various suggested techniques, it soon became evident to the MIT group that more knowledge of language and the translation process would be needed. Their most significant advances have been of a basic and fundamental nature, which will help to make it possible eventually to program computing machines to produce accurate and acceptable translations. The work that has been done on generative grammar and the theory of grammatical transformations is believed to represent an important advance in linguistics, making possible more precise descriptions of language and shedding considerable light on the relationship of syntax to some aspects of meaning. Early work, showing the necessity for sentence-for-sentence translation rather than word-for-word translation, has now met with full acceptance by all groups working in the field. The conceptual framework that the group introduced nearly 2 years ago advanced the idea that mechanical translation should be a three-step process, analysis of the incoming sentence, choice of appropriate components of the output sentence, and synthesis of the output This conceptual framework is gaining acceptance by an increasing number of mechanical translation groups. Much of the work has been concerned with the preparation of detailed grammars of English, German, and French and with continuing studies of some of the formal features of linguistic expressions, such as expressions of negation. This work is rapidly reaching fruition. The group first had to find out the best way of representing the grammar of a language for use in a machine. At the same time they have devised techniques for using the machine to aid in their research, including a programing language known as the COMIT system, for use with machines in linguistic work.

GEORGETOWN UNIVERSITY

Summary of plan of *research*

In this project it was proposed to extend the results obtained in the experiment conducted in 1954 by the institute of languages and linguistics of the university and the IBM Corp.; and to develop the additional rules required for the translation of complete texts. Research was to be concentrated initially on the analysis of contemporary Russian texts in the field of organic chemistry. An experimental approach was conceived in which groups following three different approaches would be permitted to study the problem to see which would prove the most effective. A program of computer experiments aimed at gradual improvement of one or more systems was planned. Work on other languages was planned to complement the work on Russian.

Results achieved

All three of the experimental procedures for translation of Russian were carried to the point of testing on computers. One of the methods, that of Paul Garvin, dealt primarily with the analysis of Russian syntax, while the other two methods were aimed at actual translation. It was shown that both the "code matching" technique and the "general analysis" technique had been programed to the point where crude output in English words, in which some of the problems of translation had been solved, could be demonstrated. The general analysis technique was selected for further study. Up to the present, a corpus of 268,000 running words has been utilized in the preparation of a dictionary of 10,800 entries. Furthermore, 115,000 words of Russian text have been processed by computer, and much of the output has been studied for the purpose of improving the programs. Recently, the code matching technique has been the subject of further study at the Corporation for Economic and Industrial Research at Arlington, Va., and the work of Paul Garvin has been continued at the Thompson Ramo Wooldridge Corp., Los Angeles, Calif.

As for other languages, an experimental system has been developed for French-to-English translation and has been brought to the point where French nuclear physics texts can be converted into English words which in many cases convey the thought of the original. Preliminary research has been conducted on the problems of translating English into Chinese and English into Arabic.

During the past year, all of the support of this project has come from the Central Intelligence Agency, and the work outlined above has been continued and extended to other languages.

CAMBRIDGE LANGUAGE RESEARCH UNIT

Summary of plan of research

The unit proposed to investigate the possibility of using a specially constructed mechanized thesaurus in the production of idiomatic translations by machine. To this end they planned to study the application of logic and other branches of mathematics to syntactic analysis; to extend descriptive linguistic analysis to give the cross-relations between passages in a language and translations of them into another language; and to construct comprehensive, ready-to-use mechanical dictionaries and programs for machine translation.

Results achieved

Much of the study of this group has been devoted to the semantic aspects of natural language and how to deal with them. A careful study of existing thesauri has been carried out and has served as a starting point for various experimental thesaurus-like word classification schemes which indicate the ways in which words are semantically related to each other. These classification schemes have the same form as mathematical partially ordered systems, and the unit is attempting to show that they can be so modified as to form more specialized mathematical systems known as lattices. The group believes that word schemes in lattice form will be a useful tool for natural language processing, including mechanical translation and abstracting, and information retrieval. As an example of the last type of application, a retrieval system for several hundred books has been worked out and is being expanded. Work on one particular mechanical translation scheme from Italian to English is well advanced, and other work, including construction of translation procedures based on syntactic categories, is being carried on simultaneously.

HARVARD UNIVERSITY

Summary of plan of research

It was proposed to extend the preliminary research on the structure of the Russian and English languages being carried out at the Harvard Computation Laboratory, in the light of the conviction that the processes of translation were not well enough defined to justify construction of any complete translation system. The initial effort was to be devoted to the formulation of efficient techniques for the compilation and maintenance of an automatic dictionary in order to provide an experimental tool to facilitate research still needed to develop methods for high-quality translation and a system for automatic word-by-word translation. Continued research into methods for achieving faithful, smooth translation from Russian to English was planned.

Results achieved

During the first 2 years of research, computer programs for the Univac were written which permit the operation of an automatic dictionary containing about 15,000 Russian words. The programs permit the recognition of any of the 15,000 words in any one of their forms, making it possible to process over 150,000 distinct Russian word forms. This automatic dictionary has been used to produce word-for-word translations of scientific Russian texts, which are not true translations since they fail to take account of the grammar, but which have proved useful for some purposes in lieu of actual translations. The techniques and procedures which have been developed are applicable to the whole field Of compilation and operation of automatic dictionaries.

The automatic dictionary has also been used as a tool both to compile language statistics and to conduct research on syntax and syntactic analysis. Recently developed programs for syntactic analysis are based on the work of Mrs. Ida Rhodes of the National Bureau of Standards. Programs are now in operation which provide a partial syntactic analysis of Russian sentences on an experimental basis.

UNIVERSITY OF CALIFORNIA, BERKELEY

Summary of plan of research

The purposes of the project were (1) to analyze a large amount of scientific Russian text in order to provide the information necessary for the preparation of a mechanical translation program; and (2) to write and test such a program. To minimize the size and complexity of the vocabulary, it was decided to restrict the scope of research initially to one area of science, but to design the translation system in such a way that it can be readily adapted to other fields. The more specific research tasks include compilation of an automatic dictionary and programs for its use, development of a system for the automatic parsing of sentences, and development of a mechanized system for analyzing Russian text and compiling data about the language as it is used in current scientific publications. The major part of the analysis is devoted to the solution of the "multiple meaning" problem, which will require the analysis of several hundred thousand running words of text. Programing of the translation mechanism will proceed hand in hand with the linguistic analysis, and the results of the latter will be incorporated into the program as they become available.

Results achieved

As a result of discussions with representatives of the Central Intelligence Agency and the National Science Foundation, the group decided to concentrate on the field of biochemistry, rather than nuclear physics, as originally suggested, since there already was a group studying nuclear physics.

Systems and research tools which have been produced thus far include (*a*) a maximally effective segmentation system for splitting Russian words into component parts, (*b*) a coding system for Russian grammar, (*c*) a Russian-to-English dictionary with a vocabulary coverage of over 300,000 words, (*d*) an automatic dictionary system which can accomplish look up and segmentation at a rate of 7,500 words per minute when used on an IBM 704 computer, (*e*) a system for analyzing Russian text, (*f*) an exhaustive analysis for 30,000 words of text, (*g*) a linguistic data gathering program for obtaining information from analyzed text by means of an IBM 704, (*h*) a system for coding Russian scientific text for input, (*i*) a catalog of situations in which changes in order of words are required when translating from Russian to English, and (*j*) a method for automatic parsing of Russian text.

U.S. Army

The U.S. Army has two projects in mechanical translation research.

First is a $100,000-a-year contract with the University of Texas in German-English and English-German translation under Prof. W. P.

Lehman. The Army witness testified that: "Dr. Lehman's project is a new one, less than a year old." It is a long-term research effort, based on computer analysis of parallel German and English texts. One might call this the objective approach—examine the parallel text in two languages and derive rules for translation by means of the computer. This is a tedious, systematic approach, but it has the great advantage of being equally applicable in either direction for a pair of languages. Translation of English into other languages has few "handles." By "handles" I mean identifying features built into a word which tell one immediately whether it is a noun or verb, its case, its gender, its number, and so on. We ourselves make these identifications by context—one might almost say by instinct, but really by using the vast number of tidbits of information in our minds. A machine has no such information unless every single bit has been put there, and we are very far from being able to put as many pieces of information into a machine as even a year-old child possesses. -

This is why a mathematical parallel correlation technique may prove to be the best for English to other languages and why we are supporting Dr. Lehman's work. I would like to repeat that this is a long-term project from which we cannot expect quick results.

The second Army project is with the National Bureau of Standards in Russian-English translation under Mrs. Ida Bhodes. Mrs. Rhodes' project is only slightly older than Dr. Lehman's—about a year and a half. Brig. Gen. William J. Ely, the Army witness, stated that—

> by oversimplifying, I might be to call her approach the subjective one versus Dr. Lehman's objective approach. Mrs. Rhodes is making a program for machine translation of Russian to English by a method called predictive analysis. This might be described as taking each word in the sentence and shaking it by its "handles" until it gives the machine all the possible information it contains, both about itself and about other words in the same phrase, clause, and sentence.
>
> Russian words have many "handles"—prefixes, affixes, grammatical endings, and especially word agreements. A machine can be programed to identify these "handles" and make predictions of what else must be in the same phrase, clause, or sentence. When the machine finds these predicted items, it goes on to the next problem, satisfied. If it doesn't find them, it stores the predictions and goes on with an eye cocked. If there are multiple choices, the machine makes a choice, but stores the other choices to try if the first proves wrong. Some predictions mast be fulfilled, such as having a subject and a verb, expressed or implied. Others may be fulfilled or may not. When the machine reaches the end of a sentence, it examines its hindsight pool where unfulfilled predictions are stored. If it finds any that are labeled "must be fulfilled," then it knows that the translation is probably faulty. This may happen for a number of reasons, such as printing errors, omissions, or grammatical errors by the author. This is a quick and crude explanation of Mrs. Rhodes' technique.

National Bureau of Standards

Dr. Cannon, of the National Bureau of Standards, explained Mrs. Rhodes' system in greater detail than that presented by the Army. He said:

> In our approach at the NBS the machine first examines each word in the Russian text and establishes its grammatical interpretations and meanings. I take a moment here to emphasize as strongly as possible that very rarely does a source word possess a unique interpretation, either as regards its grammatical structure or its particular shade of meaning.
>
> A simple example will suffice to illustrate this assertion. A Russian translator confronted with the isolated English word "b-o-r-e" would be hard put to decide whether it is (1) a noun describing a feature of a gun; (2) a noun characterizing a certain type of human being; or perhaps (3) the past tense of the verb

"to bear," which in itself has various meanings. We say, "The wild beast bore down upon the explorer" or "The queen bore the king a royal heir" or "The martyr bore his cross with angelic patience." In no other language, obviously, could we expect all of these connotations to be expressed also by a single word. I feel certain that one would expect similar ambiguity to exist when one goes the other way—from Russian to English. It does.

To return to our process of machine translation, the machine examines each Russian word to ascertain from its grammatical form its meaning or meanings in English, and what other Russian words it leads us to expect These expectations are pooled with others arising from the rules of conventional grammar and are compared with subsequent occurrences. Occurrences which do not match the existing predictions are stored, for further use, in what we call a hindsight pool, and are subsequently reconciled.

The English equivalents found in the dictionary during the analysis I have described are synthesized into an English sentence, giving an output which is in pidgin English and very crude, but which has the correct grammatical construction.

As far as I know, the foresight or predictive technique, which is now being called predictive analysis, together with the use of hindsight pools, originated with Mrs. Rhodes, leader of the NBS Mechanical Translation Group. I believe that it is a very powerful and a promising technique for the mechanical translation of languages, and this opinion, I am pleased to state, appears to be shared by many students of languages and members of mechanical translation groups both in this country and abroad. As you no doubt were informed yesterday, for example, Dr. Oettinger's group at Harvard University is now concentrating major effort in exploring this technique.

Other unique features of Mrs. Rhodes' approach to MT are the use of repeated passes (successive approximation techniques) to translate the more difficult sentences, and the provision, for each translated English sentence, of indicators of the measure of reliability the reader may attach to it

At the present time, her scheme is able to cope with the syntactical aspects of the mechanical translation problem. In other words, the predictions enable the machine to pinpoint the unique grammatical interpretation of a source word, so that it assumes its proper role and sequence in the target sentence. We wish we could say as much for the semantic, or multiple-meaning aspect This is a far more difficult task, as it involves the examination of the context of the word under consideration, and the derivation of the proper inference from the association of ideas revealed by the surrounding words. Remember the example, the English word "bore." Until this formidable problem is solved (and we are not yet certain that it can be) we shall be forced to print several meanings for a single-source word and, on occasion, several versions of the same sentence.

For this reason, and many others which we shall not enumerate here, our final translation will be quite inelegant—as we have said, even in pidgin English. We do feel that the crude translation yielded by our method will give the reader a correct image of the meaning in the foreign text. Consequently we are now concentrating on completing and testing thoroughly a set of computer instructions embodying our techniques.

U.S. Navy

The Navy currently supports research in machine translation of a general critical nature at Hebrew University in Israel, and at Wayne State University, where a small group is considering the translation of mathematical literature from Russian to English. This group cooperates very closely with groups at Georgetown University under CIA sponsorship, and Ramo-Wooldridge under Air Force sponsorship. The Navy is involved in a number of coordination activities in machine translation through its membership on various governmental committees. These include the CIA and NSF committees.

Following is a list of contractors of the Navy which are and have been specifically involved with research on machine translation. Those contracts concerned with other areas of data processing and information technology having a bearing, as previously mentioned, on machine translation have been omitted if they have been otherwise motivated.

Contractor	Title	Dates of contract	Amount of contract	Fiscal year 1960 funds
Massachusetts Institute of Technology, Cambridge, Mass.	High-density storage and character recognition.	April 1957 to March 1961	$96,000	$25,000
Hydel, Inc., Waltham, Mass.	High-speed photomemory.	August 1958 to August 1960.	60,000	30,000
Wayne State University, Detroit, Mich.	Russian linguistics	November 1952 to July 1956.	21,400	None
Do.....................	Mechanical translation of languages.	May 1958 to August 1960	94,500	None
United Research, Inc., Cambridge, Mass.	Print reading	June 1960 to May 1961	16,400	$16,400
International Telemeter Corp., Los Angeles, Calif.	High-capacity photomemory.	June 1953 to August 1954	28,300	None
Baird-Atomic, Inc., Cambridge, Mass.	Pattern recognition studies.	January 1954 to July 1957	100,000	None
Total funds			416,600	115,900

U.S. Air Force

Following is a list of Air Force contractors and the value of their contracts for research on mechanical translation:

Indiana University	$99, 000
Thompson Ramo Wooldridge, Inc	38, 000
University of Washington	76, 000
Syracuse University	50,000
IBM Research Center	927, 000
IBM Research Center	840, 000
Thompson Ramo Wooldridge, Inc	92,000
Baird-Atomic, Inc	381, 000
University of Milan, Italy	124, 000
National Science Foundation, for Cambridge University, England, and Harvard University	125, 000
Intelligent Machines Research Corp	267, 000

Within this broad base of contracts the Air Force started its program in automatic language translation with the search for a large capacity, high-speed device for use in lexical data handling. This search led to the glass disk storage device invented by Dr. Gilbert King. This glass disk photoscopic memory was the means to an economical and fast method of automatic language translation. This led to the program now in existence.

Figure 1 shows the dictionary in the round. Heart of the automatic language translation complex is the memory disk or dictionary—a glass word warehouse 10 inches in diameter which has as many entries as Webster's Dictionary. The 550,000 Russian-English words are stored in concentric tracks of binary code.

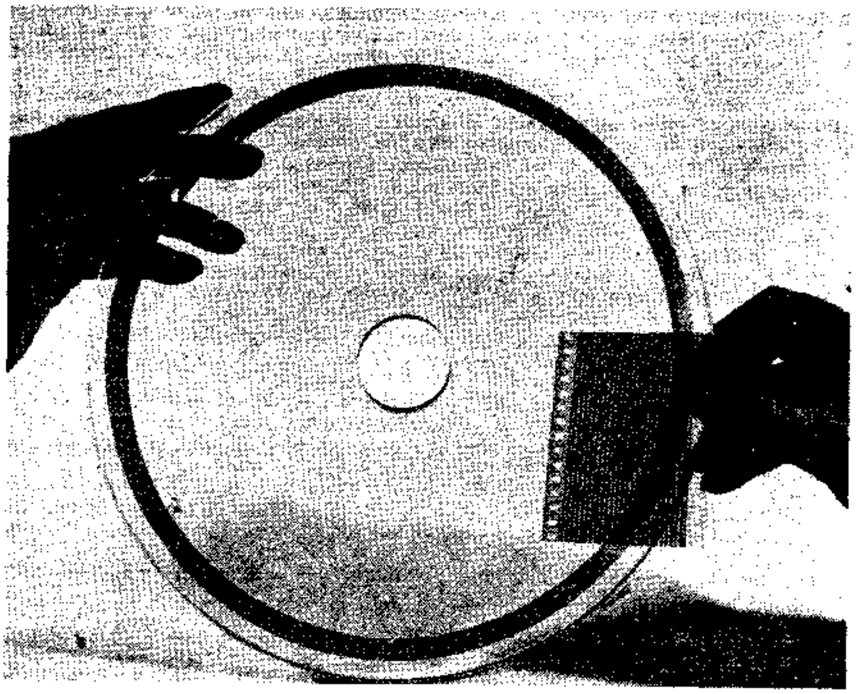


FIGURE 1

The Air Force's greatest effort is with IBM and has resulted in an experimental model of a fully automatic dictionary look-up technique. It accepts Russian word inputs, searches for the English equivalent and produces these English equivalents at the rate of approximately 30 words per second. The experimental model has been operating since last April, and has clearly demonstrated its ability to perform its basic language translation function.

An automatic print-reading R. & D. program, which will read Russian literature automatically, and at a rate comparable to the speed of the automatic dictionary, has also been supported by the Air Force. The feasibility of the print-reading technique has already been demonstrated by the contractor—Baird-Atomic, Inc.—and a more refined model will be tested by the end of this year. In describing the print-reader the Baird-Atomic witness testified as follows:

The Baird-Atomic reader, now under construction, is capable of recognizing and distinguishing a large number of different type fonts in various alphabets including at least English, Cyrillic, and Greek characters. This reader has recognition capability independent of the spacing between lines, the position of the text on the printed page, and the occurrence of randomly interspersed graphic material. Furthermore, the present design objective is to provide an instantaneous reading rate of about 1,000 characters per second. This speed is accomplished by an optical system which permits comparison of an unknown character, printed letter or number simultaneously with each of a large set of reference characters. The set of reference characters, called the system's memory, are photographic masks or optical apertures that can be rapidly and inexpensively prepared and replaced should a change in the language or type font be desirable. Baird-Atomic simultaneous comparison, recognition, and identification of characters reduces excessive mechanical motions and eliminates time-consuming electrical scanning procedures which have been proposed and utilized previously by others. Mechanical and electrical scanning procedures usually require ancillary computers of considerable complexity and cost.

The computer requirements for such a scanning reader appear to be particularly severe when Cyrillic, English, and Greek characters are intermixed as is the case of Russian technical literature. For these reasons Baird-Atomic has employed a simple optical approach to the problem.

Because of a multiplicity of related and unrelated problems associated with automatic print reading, Baird-Atomic has limited its efforts on the present contract to reading text which is first prepared on transparent 70 millimeter film. The developed film negative is read by the machine. This approach was taken for several reasons which relate principally to the manner in which technical papers are presented by their authors. They usually include pictures, graphs, and multilined mathematical equations, and to cope with these random inserts when word translation is the prime interest, would increase the cost of the development considerably. Furthermore, although the problems are not insurmountable, the availability in time of such a functional system would be questionable.

The decision to work with photographic transparent text does not, in any way, preclude the possibility of reading opaque material directly. In fact, Baird-Atomic is presently considering the elements of opaque text reading independently of the program being discussed here.

This discussion is intended to be a layman's description of the physical principles underlying the Baird-Atomic method of optical character recognition. It is presented to familiarize you with the simplicity and the ultimate potential of the technique.

Figure 2 is a photograph of a typical optical mask or array of apertures which includes characters of a particular type font.

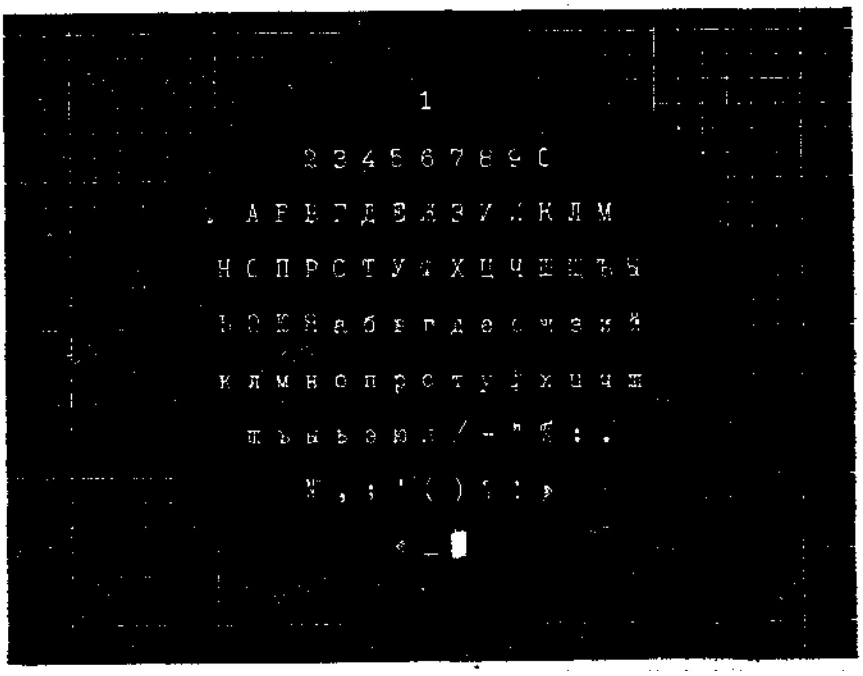


FIGURE 2

Figure 3 shows a simplified optical layout of some of the elements used to perform the character recognition. There is a source of light, a diffuser, and a lens to concentrate the diffused light onto an unknown transparent character. If opaque copy is used, this initial optical arrangement would be altered but the following discussion will remain essentially unaltered.

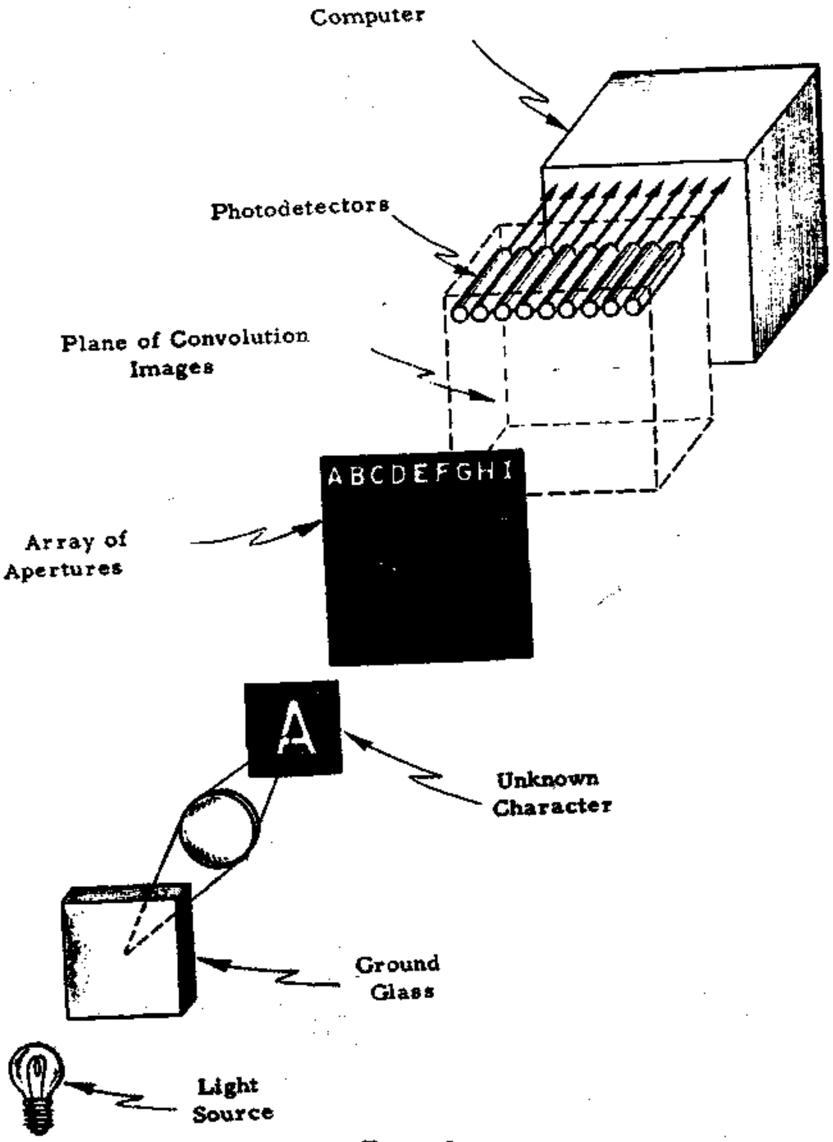


FIGURE 3

The light passing through, or reflected from, the unknown transparent character then diverges and passes through the array of standard apertures, and thence to a corresponding array of photodetectors. The electrical outputs from each of these detectors are ultimately processed and fed to the translator or to a recording tape for ultimate use with the translator.

It is noted that the mask or array of apertures includes all of the characters of interest inclusive of the one being identified. Furthermore, if the unknown character is within the master set of apertures, no motions, mechanical or electrical, are necessary to recognize or identify the unknown. Consequently, with this method of optical automatic print reading, speed is not limited by the optical recognition technique. Currently, the speed is controlled by such factors as text alignment, poor copy, and so forth.

When extended, imaginary lines, of optical rays, drawn through the center of the unknown character "A" and passing through all of the centers of all of the characters in the array of apertures pass through small holes placed in front of each of the photodetectors. Figure 4 shows a photograph of the mask and also two photographs of the distribution of light in the plane of the small holes covering each of the photodetectors. If the letter "A" is the unknown, a bright spot appears in the plane in front of the detector for recognizing "A."

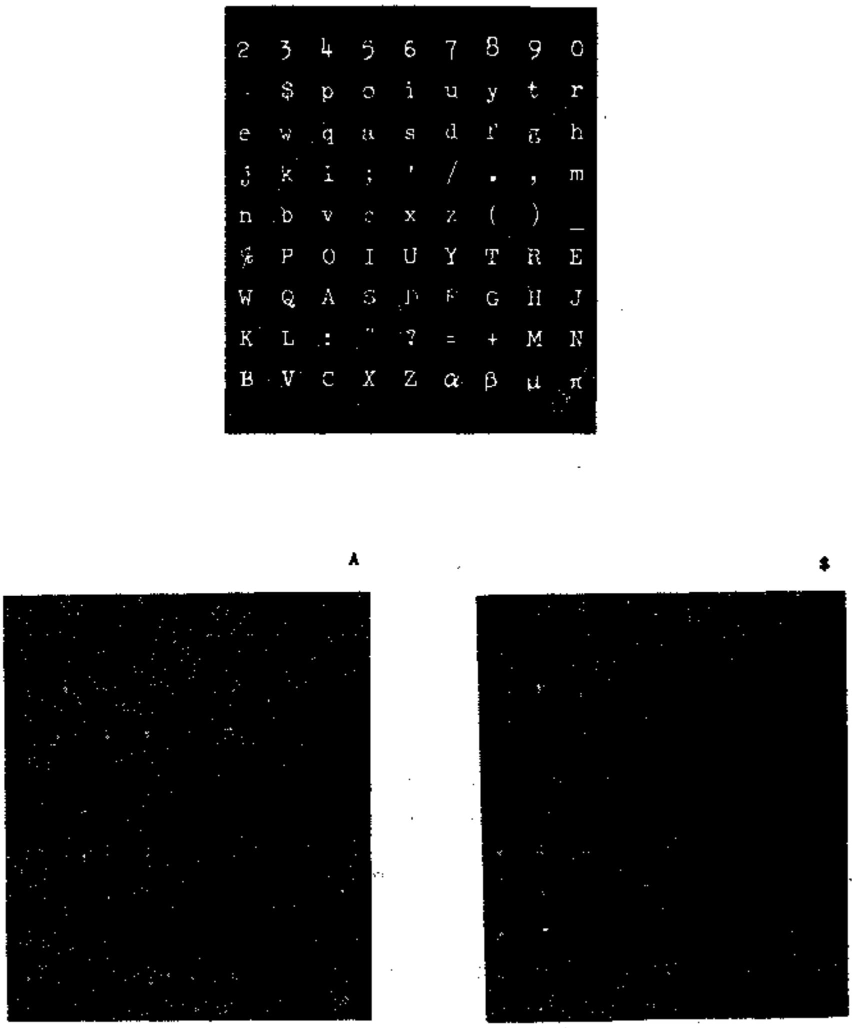


FIGURE 4

I think if we consider this page, this very last page, the upper diagram is typical of the mask, and you notice the letter "A," is three up and three over, on the mask. If you look at the lower lefthand picture, which I must say has been reproduced very poorly, there properly should be a bright dot in the position corresponding to the position of the letter "A." Actually, in the original negatives which we have, this is much more obvious than in this multilith reproduction. Similarly, if the dollar sign is used, as the unknown letter, in the next diagram one will observe a brighter spot in the second row, the second line down, indicating the identification of the dollar sign.

This method of recognizing alphanumeric characters simultaneously correlates all of the characters in a particular type font with the unknown. The decision-making function is carried out electronically by means of threshold circuits at the outputs of the photodetectors.

Furthermore, with only a slight increase in the optical complexity of the system, minor dissimilarities in characters and punctuation marks such as periods, commas, and semicolons are readily recognized. It is anticipated that appropriate identification of punctuation marks will be of primary importance in effective mechanical translating. This is believed to be a feature of this approach to optical reading which is not inherent in other disclosed techniques.

Concerning the print out of the translation in English, research is in progress for the development of a technique for automatic insertion of equations, graphs, charts, and pictures.

An integration of the above developments is expected to be accomplished and demonstrated by the fall of 1961.

Central Intelligence Agency

First in the machine translation program, the CIA is following the pragmatic empirical approach and has supported the efforts of Prof. Leon E. Dostert at Georgetown University. The combination of all Government funds to Georgetown University totals $730,000. This includes $106,600 from NSF, $120,000 from DOD, and the remaining $503,400 from CIA.

The Georgetown program has been concerned with experiments in Russian organic chemistry and French nuclear physics texts. The experiments are based on actual texts and these texts are used as the lexical and grammatical basis for the MT operations.

A total of 395,000 words in continuous texts in the field of organic chemistry in Russian have been key punched, along with a corpus of 20,000 words in the field of metallurgy, for a grand total of 415,000 key-punched running words. Their coded dictionary now includes 10,800 entries.

In the field of French to English translation, the research led to the formulation of a generalized programing system called simulated linguistic computer. This system is being used in part in the conversion of the Russian-English "Georgetown Automatic Translation Technique."

Along with the experiment-focused research, a certain amount of theoretical investigation has taken place in the following fields:

1. An approach to the establishment of semantic categories;
2. The broadening of syntactic analysis; and
3. The development of a program for the machine composition of chemical terms not found in the machine lexicon.

Prof. Leon Dostert of Georgetown University requested the committee to conduct a test of the output of the Georgetown automatic translation technique. He proposed that a random text be selected, translated, and evaluated by a qualified chemist. The results of that test are included in the appendix (p. 41) of this report.

VII. Committees and Coordination

Two committees of Government concerned with machine translation have been established.

Subcommittee on Mechanical Translation

The first committee was organized by the Central Intelligence Agency and is the Subcommittee on Mechanical Translation of the Committee on Documentation of the U.S. Intelligence Board. The CIA, therefore, had taken steps to formalize within the intelligence community, and to inform its members of new projects and of the status of existing projects. For overall coordination in matters transcending the interests of the intelligence community, CIA looks to the National Science Foundation. Membership of the Subcommittee on Mechanical Translation includes the Army, Navy, Air Force, Department of State, National Security Agency, and CIA, with a National Science Foundation representative as an associate member.

Interagency Committee on Mechanical Translation Research

The first meeting of the Interagency Committee on Mechanical Translation Research was held on March 10, 1960. This Committee was organized by the National Science Foundation following a series of informal meetings by agencies concerned with machine translation research. The Committee was organized to broaden the field of the CIA Subcommittee on Mechanical Translation. This Committee reports administratively to the Federal Advisory Committee on Scientific Information, and through that Committee all interested agencies will be kept informed of problems and progress in the field of mechanical translation research. Membership includes representatives from the Army, Navy, Air Force, CIA, and the National Science Foundation.

Coordination

The NSF pointed out the coordination of the overall program through the existing committees, the work of the NSF in managing grants for other Government agencies, their coordinating effort, and their directional control. This has not been a program within the Federal Government where each agency was working in isolation. The NSF witness stated that there has been close cooperation and coordination in the support of research on this very difficult problem.

The CIA witness testified that it is imperative that an agency like the NSF be in a coordinating position.

The Air Force witness, in his prepared statement, said an important element in the Air Force program is participation in other automatic language translation research efforts. He maintains direct cooperation with the National Science Foundation and with the Central Intelligence Agency. The Air Force also keeps abreast with automatic language translation developments of the Army, the Navy, the National Bureau of Standards, as well as with all other research efforts throughout the United States.

It was also noted that there are a number of national and international symposia on the subject of mechanical translation research sponsored by various Government agencies.

VIII. Status of Human Translation of Russian

In his presentation, Mr. Paul Borel, of CIA, mentioned that the Russian scientific output of scientific information is now about 780 million words a year. He also stated that the United States was now translating into English 53 million words a year, or about 7 percent of the total output.

Following questioning from the committee, Mr. John J. Bagnall, of the CIA, added that the CIA scans half of the 780 million words that are made available, and that through other agencies of Government virtually all of the available Russian literature is scanned.

The major value of the Russian work is in the periodical literature. The largest part of the 780 million words is contained in the monographs or books in scientific subjects. The CIA finds that ordinarily the books contain information in scientific fields which has been previously published in the periodical literature, and is subsequently collated in book form appearing several years later. Consequently, by scanning the periodical literature which is current and up to date, much of the book literature may be left out.

Mr. Bagnall further testified that—

> the CIA is aware of and scans the literature made available. For complete analysis of much of this literature, full translation would be required. Full translations take considerable time, as you well realize, and of course if the translations could be made available much more rapidly for analysis in connection with current events and developments, it would be advantageous.
>
> All of the information, shall we say extracted or translated from the Russian scientific literature, is available to other Government agencies and to the general public. For example, the particular output of CIA is a scientific information report summarizing the highlights of developments in the Russian scientific literature, and is issued by the Office of Technical Services, Department of Commerce, on subscription to the general public.

In commenting on the timelag when the Russian text reaches this country and the time it is abstracted or translated and made available to the, scientific community, Mr. Bagnall commented that—

> in general, the highlights of important items may be abstracted and made available within 4 to 6 weeks after date of publication. However, translations of important articles will be made available, depending on their length, of course, at considerably later dates, running from 1 month to perhaps 6 months for a book of, say three or four hundred pages.

Mr. Bagnall further stated that the Russians have an enormous organization in Moscow for translation and abstracting. There are some 2,000 translators and 20,000 abstractors used part or full time for just our scientific and technical literature.

In addition to the statements from the CIA witness, Dr. King, of I.B.M., submitted the following chart from the Planning Research Corp. study, showing the delays in human translations (fig. 5).

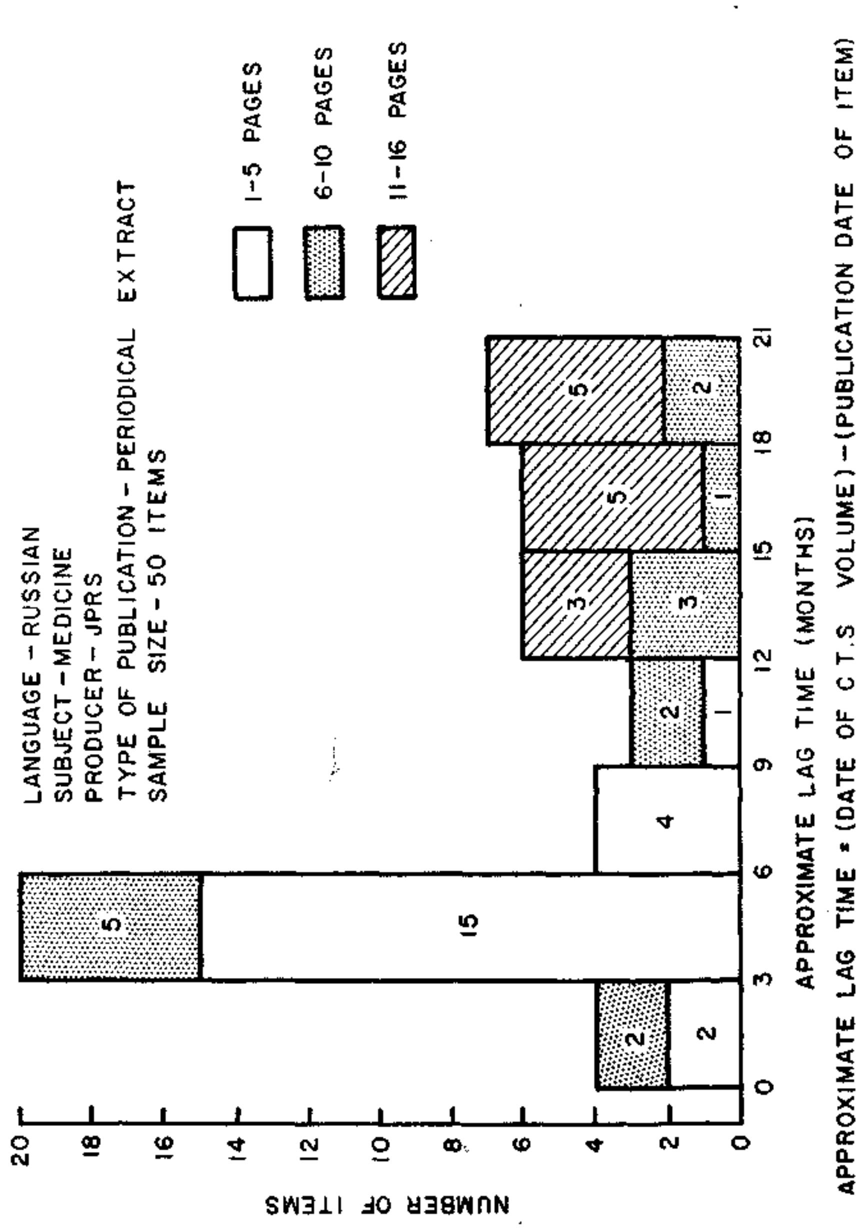

LANGUAGE - RUSSIAN
SUBJECT - MEDICINE
PRODUCER - JPRS
TYPE OF PUBLICATION - PERIODICAL EXTRACT
SAMPLE SIZE - 50 ITEMS
1-5 PAGES
6-10 PAGES
11-16 PAGES
NUMBER OF ITEMS
APPROXIMATE LAG TIME (MONTHS)
APPROXIMATE LAG TIME = (DATE OF C.T.S VOLUME) - (PUBLICATION DATE OF ITEM)

FIGURE 5

IX. FUTURE

Timing the arrival of an operational machine translation system is very difficult, according to the majority of the witnesses. There are no operational systems at this time, and the only optimistic estimate is for some time in 1961. However, there are doubts about the value of the output, and even though an operational complex will exist, some will continue to degrade its value unless something very close to 100 percent accuracy is achieved.

However, it is not too soon to start planning for this eventuality, and the committee investigated the future use of an operational system with numerous witnesses. Of concern to the committee was the possibility of forming a new agency of Government, or a central agency assigned to an existing branch of Government, to handle the national job of machine translating all documents at one center. Responses varied, as indicated in the following excerpts of testimony.

Dr. Adkinson, of the National Science Foundation, stated:

> It is my view that we should proceed the same as we are doing with human translators; that there will be agreement reached as to the responsibilities among not only the Government agencies, but also among private organizations interested in translating with machines if they become useful. I think we will have to come to an agreement on responsibilities. But I wouldn't envision that at first with the machines we would translate everything, because the scientists and engineers and the administrators are having enough trouble now reading. I hope they will translate the important things. This will vary, depending on whether you are an intelligence agency; whether you are a military agency without intelligence responsibility, say with research and development responsibility; whether you are the Department of Agriculture, or whether you are the Department of the Interior. So I think there will have to be an agreement on areas the same as there is agreement on areas today and a central file on what is being translated and what has been translated.

Brigadier General Ely, of the Army, did not agree with the centralized agency concept. He said—

> I don't feel that it should be concentrated, because it involves the use of general purpose computers—all types of computers, which will be available in many places.

Mr. Borel, of CIA, took the following position in his prepared statement:

> More basic is the problem of organization. Shall an MT capability once achieved be exploited by each on his own or should a central facility serve all? If the latter, who shall set it up, who shall operate it, and tinder what terms shall Government and private interests participate?
>
> It is not too early to start thinking about this. I believe a central facility is indicated, but not exclusively so. The enormous potential output of MT greatly exceeds the present and prospective requirements of any one part of Government or single private organization. Problems of procuring and selecting materials to be translated, and of disseminating translations to those needing them, are very considerable. These can most efficiently and economically be solved centrally. Moreover, a central facility permits the use of equipment exclusively designed to produce automatic translations. There are, however, requirements for accomplishing translations under mobile conditions, or, for fully utilizing general-purpose equipment acquired for processing data rather than language. Hence there is also continuing need for research to develop MT materials and programs in various languages and disciplines for translation by general-purpose computers.

Dr. King, of I.B.M. proposed a modification of the central facility. He said:

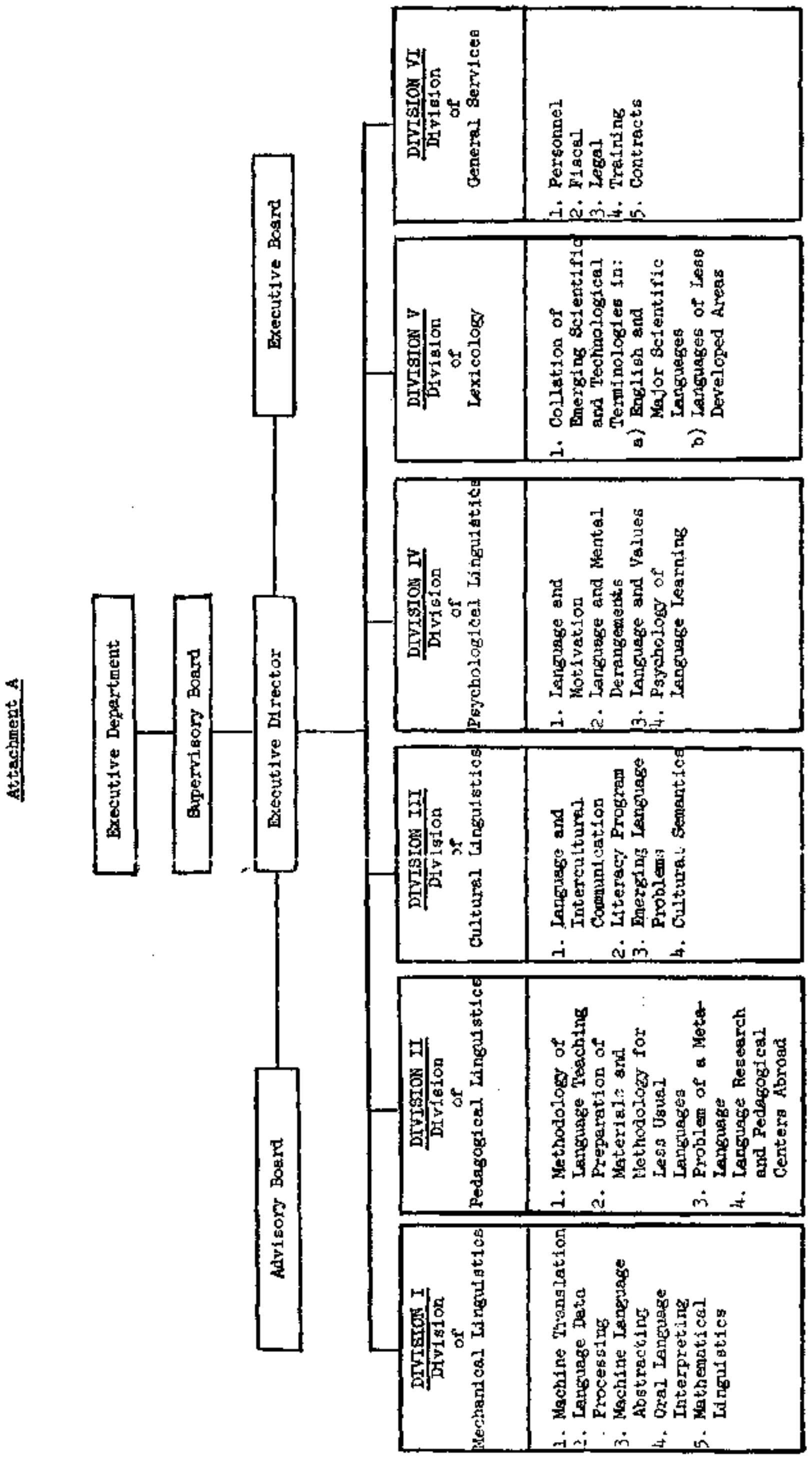
> I believe the technical part of the problem, say dictionary lookup, is well enough in hand so that one could consider a single center. It would save a lot of money if all the efforts on the linguistics work were put into this single machine organization. I think this single center could handle all the requirements and coordinate them and consolidate them for general information and intelligence for this country. I am sure that there always will be necessity to have outlying translation equipment, say with the Army for field data.

In his prepared statement for the record, Prof. W. P. Lehman, of the University of Texas, stated—

> when successful, machine translations need be made at only one center. For, if a Chinese article on the mineral resources of Tibet is once translated, it will not need further translation. Accordingly, machine translation should be arranged under an agency of the Government, such as the Department of Defense, the Library of Congress, the National Science Foundation, the Department of Health, Education, and Welfare, or a separate foundation devoted to linguistics. Such an organization might circulate all translated materials gratis, or for a fee.

When asked about the overall field of language sciences, Professor Dostert, of Georgetown University, stated:

> * * * I believe the time is at hand to study the advisability of establishing a national institute of language science, somewhat along the lines of the present National Institutes of Health. I would envisage that such an institute would embrace several basic areas: Mechanical linguistics (including machine translation), pedagogical linguistics, cultural linguistics, psychological linguistics, and lexicography. An outline and table of organization is attached.


Attachment A
Executive Department
Supervisory Board
Executive Director
Advisory Board
Executive Board
DIVISION I
Division
of
Mechanical Linguistics
1. Machine Translation
2. Language Data Processing
3. Machine Language Abstracting
4. Oral Language Interpreting
5. Mathematical Linguistics
DIVISION II
Division
of
Pedagogical Linguistics
1. Methodology of Language Teaching
2. Preparation of Materials and Methodology for Less Usual Languages
3. Problem of a Meta-Language
4. Language Research and Pedagogical Centers Abroad
DIVISION III
Division
of
Cultural Linguistics
1. Language and Intercultural Communication
2. Literacy Program
3. Emerging Language Problems
4. Cultural Semantics
DIVISION IV
Division
of
Psychological Linguistics
1. Language and Motivation
2. Language and Mental Derangements
3. Language and Values
4. Psychology of Language Learning
DIVISION V
Division
of
Lexicology
1. Collation of Emerging Scientific and Technological Terminologies in:
a) English and Major Scientific Languages
b) Languages of Less Developed Areas
DIVISION VI
Division
of
General Services
1. Personnel
2. Fiscal
3. Legal
4. Training
5. Contracts

Summary Explanation of Organization Chart

1. The NALS would have as its primary mission the coordination of the national effort in the various areas of the science of language, both practical and theoretical, on the basis of the five substantive divisions given in the chart, plus one service division.
2. In certain fields it would have an operating responsibility whenever this procedure would be deemed more effective than the assignment of specific research and development projects to selected academic institutions.
3. It would provide for interfellowships to permit the assignment for 1 year or more of top specialists who would be working with the NALS in specific research or development projects.
4. The Advisory Board would be made up of five recognized scholars or authorities in the five substantive fields who would function as regular consultants.
5. The Executive Board would be made up of the divisional directors under the chairmanship of the Executive Director.
6. A Supervisory Board could be made up of five executives from academic institutions or learned associations who would be appointed on a rotating basis. The Supervisory Board would also include two members from the Government.
7. It is estimated that when in full operation the NALS would have a staff of from 300 to 500 persons. The preliminary budgetary estimate is of the order of $15 million per annum, half of which would be for internal operations, and the remainder for external projects, based on contractual grants.
8. The first measure to be taken if the project meets with approval would be the appointment of a small planning staff under the direction of the future Executive Director, who would review and present a more complete and rigorous proposal for legislative consideration. The planning staff would arrange to have the advice of five recognized authorities specialized in diverse areas within the purview of the proposed Academy, who presumably would become the future Advisory Board.

X. CONCLUSIONS

1. The hearings on mechanical translation vividly pointed out the importance of a mechanical translation system to the overall intelligence and scientific effort of our Nation. With the advent of such a capability, a new approach will be taken by all segments of our culture to the reading of foreign documents. Truly, a capability of translation in reverse—that is, English into foreign languages—will open up new vistas and avenues for the exchange of cultural, economic, agricultural, technical, and scientific documents that will present the American way of life to people throughout the world.

The pursuit of this research and development program on mechanical translation is a must and should be vigorously continued to insure an early capability on a national effort.

2. There are a number of approaches being taken to solve the research problem. These were listed as statistical, systematic, empirical, objective, and subjective. The definitions of these approaches may be partially overlapping. All approaches are valid and should be pursued so that the Nation will benefit from an early interim capability while waiting for the long-term research to provide a highly accurate system.

3. The equipment oriented program of the Air Force provides a unique tool for the checking out of the complex fully automated system of input, logical processing, and output. This program provides for the inclusion of modifications derived from other research as the results are made known to the overall mechanical translation family. It is now time for the mechanical translation research community to start combining the good features of the several programs and come up with the best workable program for the first application.

4. There are many Government agencies conducting research in this field. With so many agencies working in this field it will eventually make coordination of the total effort more complicated, and lead to the possibility of duplication in some areas.

5. The National science Foundation, although participating in this research program on mechanical translation, is not assuming the predominant leadership role that is expected of the Foundation. Greater effort could be made by the National Science Foundation in establishing a coordinated program, managing more of the diversified programs, and generally directing the national effort.

6. The Georgetown University proposal for a National Academy of Language Sciences along the line of the National Institutes of Health has considerable merit. This is indeed a steppingstone toward the solution of the language barriers that confront our Nation in international relations. One segment of such an academy would be devoted to the national program of mechanical translation.

7. The most promising long-range program for mechanical translation is Army-financed research at the National Bureau of Standards. There was almost complete agreement that this method of syntactical analysis would produce the best readable translation as far as sentence structure is concerned.

8. In the Department of Defense all three services have their own programs. The effort in the DOD could well be centered in one service, with the requirements of the other services met by the responsible agency.

9. There are apparently sufficient funds allocated by all Government agencies to maintain a satisfactory rate of progress in this field of research.

10. A national center for machine translation will be required in the near future. The national center could well have branches in specific areas throughout the world, but the center would be the overall coordinating agency for translation. Exceptions for intelligence and military necessity will undoubtedly be made.

11. An eventual national machine translation production program will most likely be operating on a full-time basis. A special-purpose computer, designed for translation and not including the other costly benefits of an all-purpose computer, will be desirable.

12. An early production program, on a limited scale, as proposed by Georgetown, is an excellent approach. The determination, at an early date, of a limited machine translation with postediting will provide the scientific community with a sample of things to come. The value cannot be assessed as there are insufficient data available. This in turn justifies the limited production so that comments can be received from the users and the value of such a limited capability subsequently assessed.

13. The committee is very pleased with the present effort of manual translation of Russian documents, periodicals, and books now being accomplished by the CIA and other Government agencies. The CIA is to be congratulated for such an outstanding contribution to this field of intelligence collection.

APPENDIX

Results of the Georgetown University test: This test was requested to prove the value of the present method of the Georgetown automatic translation technique. A random text was selected in the field of chemistry (1), the text was translated by machine (2), and evaluated by Dr. Francis Weiss. Dr. Weiss' evaluation was submitted to the committee by the Chief of the Science and Technology Division of the Library of Congress (3). A human translation is included for cross-reference (4).

RANDOM TEXT (1)

ИССЛЕДОВАНИЯ ИЗ ОБЛАСТИ ХИНОЛИНА И ЕГО ПРОИЗВОДНЫХ

XXI. СОВМЕСТНАЯ КОНДЕНСАЦИЯ АРИЛАМИНОВ С ГИДРАКРИЛОВЫМ АЛЬДЕГИДОМ

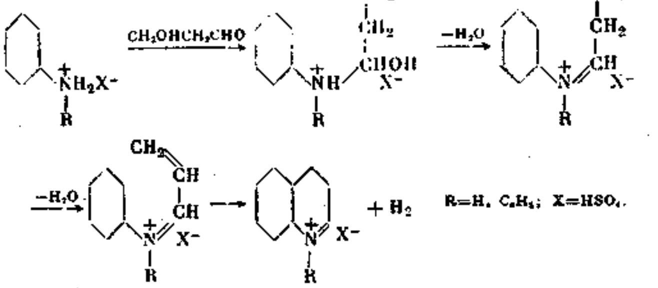
Б. И. Ардашев и *В. И. Минкин*

Ранее было отмечено, что гидракриловый (β-оксипропионовый) альдегид может принимать участие наряду с акролеином в реакции Скраупа [1]. В самом деле, он образуется в значительных количествах как при разложении глицерина [2], так и при присоединении к акролеину воды при повышенной температуре [3], причем реакция присоединения катализируется кислотой [4]. В связи с этим представлялось интересным изучить взаимодействие некоторых ароматических аминов с гидракриловым альдегидом с целью получения хинолинов. Эта реакция не описана в химической литературе, за исключением патента Чичибабина [5].

Нами было установлено, что уже в обычных условиях проведения синтеза Скраупа, а именно — при прикапывании гидракрилового альдегида в нагретую до 120°—140° реакционную смесь получается хинолин с выходом 15%. Если применять более мягкие условия - использовать окислитель *м*-нитробензолсульфокислоту и вести реакцию в разбавленном растворе, выходы хинолинов при использовании гидракрилового альдегида достигают 50%. Это позволяет заключить, что при обычном проведении реакции Скраупа с глицерином, а также при замене глицерина акролеином [[4]] одним из направлений реакции является взаимодействие с гидракраловым альдегидом, образующимся из глицерина или из акролеина.

Реакция гидракрилового альдегида с ароматическими аминами распространяется на диариламины. Так, из дифениламина нам удалось получить соль N-фенилхинолиния, и эта реакция применена к другим диариламинам. Таким образом, показана возможность введения вторичных аминов в модификацию синтеза Скраупа с гидракриловым альдегидом.

Из нескольких направлений реакции взаимодействия гидракрилового альдегида с ариламинами, ведущих к образованию хинолинов, наиболее вероятной представляется нам схема, согласно которой из анила гидракрилового альдегида возникает анил акролеина, циклизующийся в хинолин.



Литературные источники, относящиеся к изучению механизма реакции Скраупа,* и полученные нами экспериментальные данные в этой работе приводят нас к выводу, что реакция Скраупа протекает как через взаимодействие с акролеином, так и с гидракриловым и, вероятно, с глицериновым альдегидом,** причем в зависимости от исходных продуктов и принятых условий какое-либо из этих направлений является главным.

ЭКСПЕРИМЕНТАЛЬНАЯ ЧАСТЬ

Реакции проводились в трехгорлой колбе, снабженной механической мешалкой, обратным холодильником, в который вставлялась капельная воронка, и термометром. Вначале опыты проводились с самим гидракриловым альдегидом; однако затем ввиду полной идентичности получающихся результатов применялся диэтиловый ацеталь его этилового эфира как более легко доступный [5]. Чтобы избежать излишнего осмоления и снижения выходов, альдегидные компоненты всегда дважды ректифицировались, и использовался продукт с т. кип. 91—92° при 32 мм.

Опыт 1. Хинолин. К 25 мл H_2SO_4 (*d* 1.84) и 12.5 г нитробензола прибавлялись 9.3 г анилина. Смесь нагревалась до 120°, и в течение 0.5 часа прибавлялись по каплям 17 г диэтилового ацеталя β-этоксипропионового альдегида, причем температура не поднималась выше 130°. После этого смесь кипятилась еще 1—2 часа при 138—148°. Хинолин

* Например, указание Кана, что при аналогичном использовании акролеина получаются лишь следы хинолина [7].

** Это направление реакции Скраупа нами исследуется.

выделялся желтой кровяной солью [8]. Выход 2.05 г (15.6%). Т. кип. 234 240°; пикрат т. пл. 201°.

Опыт 2. Хинолин. Раствор *м*-нитробензолсульфокислоты был приготовлен, по Утермолену [9], из 12.5 г нитробензола, 50 г 20% олеума и 20 мл воды. К нему добавлялись 9.3 анилина; затем при 125—135° в течение 30—45 минут прикапывались 20 г диэтилацеталя β-этоксипропионового альдегида, после чего смесь нагревалась при 135—145° еще 1—1.5 часа. Выход хинолина 6.95 г (53.8%). Т. кип. 233—238°; пикрат т. пл. 200°.

Опыт 3. 6-Метоксихинолин. Опыт аналогичен опыту 2. Из 12.3 г *п*-анизидина получено 6.74 г (43.4%) 6-метоксихинолина. Т. кип. 278—284°; пикрат т. пл. 212°.

Опыт 4. Перхлорат N-фенилхинолиния. К смеси 8.5 г дифениламина, 45 г нитробензола и 10 мл H_2SO_4 за 15 минут прибавлено при 120—130° 10 г диэтилацеталя β-этоксипропионового альдегида, и смесь нагревалась еще 2 часа при 125—135°. Выделение проводилось аналогично изолированию N-арилхинальдиниевых солей [10]. Выход перхлората 1.72 г (11.2%). После перекристаллизации из воды т. пл. 157°.

Найдено %: С 58.75; Н 3.79; Cl 11.83; ClO_4 33.78. $C_{15}H_{12}NClO_4$. Вычислено %: С 58.93; Н 3.95; Cl 11.60; ClO_4 32.69.

Выводы

1. Реакцией гидракрилового альдегида с анилинами получены хинолиновые производные.
2. Показана возможность введения в реакцию Скраупа диариламинов. Синтезирован перхлорат N-фенилхинолиния.
3. Обсужден механизм и предложена схема реакции.

ЛИТЕРАТУРА

[1] Б. И. Ардашев, ЖОХ, *16*, 47 (1946). — [2] Р. Отшильд, И. Ита, Химия и химич. технол., *6* (91), 40 (1957). — [3] A. Nef, Lieb. Ann., *335*, 219 (1905). — [4] Ф. Пинтерт, Синтезы орг. преп., ИЛ, *3*, 88 (1952). — [5] А. Е. Чичибабин, фр. пат. 981466, 28 V 1951; Ch. A., *49*, 383 (1955). — [6] Б. И. Ардашев, Усп. хим., *23*, 46 (1954). — [7] H. L. Jale, J. Bernstein, J. Am. Chem. Soc., *70*, 254 (1948). — [8] Н. С. Козлов, ЖОХ, 7, 1864 (1937). — [9] W. P. Utermohlen, J. org. Ch., *8*, 544 (1943). — [10] В. И. Минкин, Б. И. Ардашев, ЖОХ, *28*, 2556 (1958).

Поступило в Редакцию
21 ноября 1957 г.

Новочеркасский политехнический институт.

MACHINE TRANSLATION (2)

1

AN INVESTIGATION FROM A RANGE XINOLINA AND ITS DERIVATIVE 200 00009
RRXXI. COMMON CONDENSATION ARILAMINOV S GIDRAKRILOVYM BY ALDEHYDE 00017
B. I. ARDAWEV AND V. I. MINKIN 00025
IT PREVIOUSLY WAS NOTED, THAT GIDRAKRILOVYI /G2-OKSIPROPIONOVYI/ ALDEHYDE 00036
CAN ASSUME PARTICIPATION ALONG S AKROLEINOM IN A REACTION SKRAUPA //1/. IN 00049
FACT, IT IS FORMED IN SIGNIFICANT AMOUNTS BOTH ON THE DECOMPOSITION OF 00061
GLYCERIN //2/, AND UPON ADDITION TO AKROLEINU WATER AT TO THE INCREASED 00075
TEMPERATURE //3/, .. THE REACTION OF ADDITION CATALYZES BY ACID //4/. IN A 00090
BOND WITH THAT THERE PRESENTED INTERESTING TO STUDY THE INTERACTION OF CERTAIN 00098
AROMATIC AMINES S GIDRAKRILOVYM BY ALDEHYDE WITH THE PURPOSE OF OBTAINING 00106
XINOLINOV. THIS REACTION NOT WAS NOT DESCRIBED IN THE CHEMICAL LITERATURE. 00116
FOR THE EXCLUSION OF THE PATENT OF CHICHIBABIN //5/. 00124
BY US WAS ESTABLISHED, THAT ALREADY IN THE USUAL CONDITIONS OF THE 00134
CONDUCTING OF SYNTHESIS SKRAUPA, BUT PARTICULARLY - UPON ADDITION 00143
GIDRAKRILOVOGO ALDEHYDE INTO HEATED UP TO 120# - 140# A REACTION MIXTURE IS 00155
OBTAINN XINOLIN WITH A YIELD 15%. YIELDS XINOLINOV BY UTILIZATION 00182
GIDRAKRILOVOGO ALDEHYDE AMOUNT TO 50%. THIS PERMITS CONCLUDING, THAT UPON THE 201 00195
USUAL CONDUCTING OF A REACTION SKRAUPA WITH GLYCERIN, BUT ALSO UPON THE 00205
SUBSTITUTION OF GLYCERIN AKROLEINOM //6/ ONE FROM THE DIRECTIONS OF A REACTION 00215
THERE IS INTERACTION S GIDRAKRILOVYM BY AN ALDEHYDE, WHICH ARE FORMED FROM 00223
GLYCERIN OR FROM AKROLEINA. 00228
A REACTION GIDRAKRILOVOGO ALDEHYDE WITH AROMATIC AMINES EXTENDS TO 00237
DIARILAMINY. THUS, FROM DIFENILAMINA TO US WAS POSSIBLE TO OBTAIN SALT 00247
N-FENILXINOLINI4, AND THIS REACTION USED TO OTHER DIARILAMINAM. THUS, IT IS 00260
SHOWED THE POSSIBILITY OF THE INTRODUCTION OF SECONDARY AMINES INTO THE 00266
MODIFICATION OF SYNTHESIS SKRAUPA S GIDRAKRILOVYM BY ALDEHYDE. 00273
FROM SEVERAL THE DIRECTIONS OF THE REACTION OF INTERACTION GIDRAKRILOVOGO 00280

2

ALDEHYDE S ARILAMINAMI, LEADING TO A FORMATION XINOLINOV, MOST PROBABLE THERE 00291
AM REPRESENT TO US A PLAN, ACCORDING TO WHICH FROM ANILA GIDRAKRILOVOGO 00300
ALDEHYDE IT ARISES ANIL AKROLEINA, GIKLIZUH5II IN XINOLIN. 00309
00310
/FORMULA/ 00310
00310
00311
LITERATURE SOURCES, WHICH ARE RELATED TO THE STUDY OF THE MECHANISM OF A 00319
REACTION SKRAUPA, AND OBTAINN BY US EXPERIMENTAL GIVEN IN TO THIS WORK LEAD US 00332
TO A CONCLUSION, THAT A REACTION SKRAUPA PROCEEDS BOTH THROUGH INTERACTION S 00343
AKROLEINOM, AND S GIDRAKRILOVYM AND, PROBABLY WITH A GLYCERIN ALDEHYDE, ., IN 00357
RELATION TO FROM INITIAL PRODUCTS AND ASSUMED CONDITIONS KAKOE-LIBO FROM THIS 00368
DIRECTIONS THERE IS MAIN. SKRAUPA FTNB FOR EXAMPLE, INFORMATION I3L4, THAT 00380
UPON ANALOGOUS UTILIZATION AKROLEINA THERE ARE OBTAINN ONLY TRACES XINOLINA 00389
//7/. FTNE BY ALDEHYDE FTNB THIS DIRECTION OF A REACTION SKRAUPA BY US IS 00400
STUDIED. FTNE 00403
AN EXPERIMENTAL PART 00406
A REACTION WERE CONDUCTED IN A THREE-NECKED FLASK, WHICH WAS EQUIPPED BY 00414
A MECHANICAL STIRRER, A REFLUX CONDENSER, INTO WHICH THERE WAS INSERTED A 00423
DROPPING FUNNEL, AND A THERMOMETER. AT FIRST EXPERIMENTS WERE CONDUCTED S 00433
SAMIM GIDRAKRILOVYM BY ALDEHYDE., HOWEVER THEN IN VIEW OF FULL IDENTICNOSTI 00442
WHICH ARE OBTAINN RESULTS THERE WAS USED DIETHYL AQETAL6 ITS ETHYL ESTER AS 00451
MORE EASILY ACCESSIBLE //5/. IN ORDER IZBEJAT6 IZLIWNEGO TARRING AND 00463
REDUCTIONS OF YIELDS, AL6DEGIDNYE COMPONENTS ALWAYS TWICE REKTIFIQIROVALIS6, 00472
AND WERE USED A PRODUCT WITH T. BOILING 91 - 92@ AT 32 MM.. 00486
AN EXPERIMENT 1. XINOLIN. TO 25 ML. H#2SO#4 /SD 1.84/ AND 12.5 G OF 00502
NITROBENZENE THERE WERE ADDED 9.3 G OF ANILINE. A MIXTURE WAS HEATED UP TO 00511

3

120@, EVEN DURING A0.5 HOUR WERE ADDED DROPWISE 17 G DIETHYL AQETAL4 00526
G2-3TOKSIPROPIONOVOGO ALDEHYDE, ., A TEMPERATURE DID NOT RISE ABOVE 130@. 00537
AFTER THIS A MIXTURE WAS BOILED MORE 1 - 2 A HOUR AT 138 - 148@. XINOLIN WAS 00553
ISOLATED YELLOW BLOOD BY SALT //8/. A YIELD 2.05 G /15.6%/. T. BOILING 234 - 202 00574
240@., PIKRAT T. M. 201@. 00583
AN EXPERIMENT 2. XINOLIN. A SOLUTION M-NITROBENZOLSUL6FOKISLOTY WAS 00591
PREPARED, BY UTERMOLENU //9/, FROM 12.5 G OF NITROBENZENE, 50 G 20% OLEUMA AND 00610
20 ML. OF WATER. TO IT THERE WERE ADDED 9.3 OF ANILINE., THEN AT 125 - 135@ 00626
DURING 30 - 45 MINUTES THERE WERE ADDED 20 G DI3TILAQETAL4 00636
G2-3TOKSIPROPIONOVOGO ALDEHYDE, AFTER WHICH A MIXTURE WAS HEATED AT 135 - 145@ 00648
MORE 1 - 1.5 A HOUR. A YIELD XINOLINA 6.95 G /53.8%/. T. BOILING 233 - 238@ 00669
. PIKRAT T. M. 200@. 00676
AN EXPERIMENT 3. 6-METOKSIXINOLIN. AN EXPERIMENT ANALOGICEN TO AN 00683
EXPERIMENT 2. FROM 12.3 G $N-ANIZIDINA WERE OBTAINN 6.74 G /43.4%/ 00697
6-METOKSIXINOLINA. T. BOILING 278 - 284@., PIKRAT T. M. 212@. 00712
AN EXPERIMENT YA. PERXLORAT N-FENILXINOLINI4. TO A MIXTURE 8.5 G 00722
DIFENILAMINA, 45 G OF NITROBENZENE AND 10 ML. H#2SO#4 FOR 15 OF MINUTES ADDED 00735
AT 120 - 130@ 10 G DI3TILAQETAL4 G2-3TOKSIPROPIONOVOGO ALDEHYDE. AND A MIXTURE 00748
WAS HEATED MORE A 2 HOUR AT 125 - 135@. EVOLUTION WAS CONDUCTED SIMILARLY 00761
IZOLIROVANIH N-ARILXINAL6DINIEVYX SALTS //1/0/. A YIELD PERXLORATA 1.72 G 00773
/11.2%/. AFTER RECRYSTALLIZATION FROM WATER T. M. 157@. 00786
IT IS FOUND%.. C 58.75., H IT IS., C$L IT IS., C$LO#4 IT IS. 00802
C#1#5H#1#2NC$LO#4. IT IS CALCULATED%.. C 58.93., H IT IS., C$L IT IS., C$LO#4 00817
IT IS. 00819
CONCLUSIONS 00821
1. BY A REACTION GIDRAKRILOVOGO ALDEHYDE WITH ANILINES WERE OBTAINN 00829
XINOLINOVYE DERIVATIVE. 00832

4

2. WERE SHOWED THE POSSIBILITY OF INTRODUCTION INTO A REACTION SKRAUPA 00840
DIARILAMINOV. IT IS SYNTHESIZED PERXLORAT N-FENILXINOLINI4. 00846
3. WERE DISCUSSED A MECHANISM AND WAS PROPOSED A PLAN OF A REACTION. 00855
THE LITERATURE 00857

/1/ B. I. ARDAWEV, J.GEN.CHEM., 16, 47 /1946/. - /2/ R. OTWIL6D, P. PTI. 00884
CHEMISTRY AND XIMIC. TEXNOL., 8 /91/. 40 /1957/. - /3/ A. NEF, LIEB. ANN., 00909
335, 219 /1905/. - /4/ F. PINTERT, SYNTHESES ORG. PREP., IL, 34 88 /1952/. - 00938
/5/ A. E. CHICHIBABIN, FR. PAT. 981466, 28 IN 1951., CH. A., 49, 383 /1955/. 00962
- /6/ B. I. ARDAWEV, USP. CHEM., 23, 46 /1954/. - /7/ YU. LENINGRAD JALE, ZH. 00989
BERNSTEIN. - /8/ N. S. KOZLOV, J.GEN.CHEM., 7, 1864 /1937/. - /9/ W. P. 01026
UTERMOHLEN, ZH. ORG. CH., 8, 544 /1943/. - /10/ V. I. MINKIN, B. I. ARDAWEV, 01060

RECEIVED 21 NO4BR4 1957 G. 01069

NOVOCERKASSKII POLITEXNICESKII A INSTITUTE. 01074

01074

EVALUATION (3)

THE LIBRARY OF CONGRESS,
SCIENCE AND TECHNOLOGY DIVISION,
Washington, D.C., June 1, 1960.

Col. FRANK DILLON,
House Science and Astronautics Committee,
Washington, D.C.

DEAR COLONEL DILLON: I have asked Dr. Francis J. Weiss, a member of this Division, to review the machine translation of the Russian chemistry item in which you are interested. I have attempted to summarize here his rather lengthy report to me which commented both on the content of the item as well as the quality of the translation. We will be happy to provide an extended review if you desire. Please feel free to call Dr. Weiss on code 173, extension 597.

In brief, the translation can be understood by a chemist thoroughly familiar with the subject under discussion. A competent chemist can, from a study of the chemical formulas, reconstruct many of the chemical terms which appear to be a cross between a translation and a tranliteration. With this background in chemistry, it is possible to work ahead slowly and systematically in order to make up for the obvious linguistic deficiencies of the machine translation and to grasp the principal train of thought. The reviewer was able to ascertain readily the essence of the paper which is contained in the paragraph running from 00134 to 00228.

The translation of that part of the paper dealing with the experiments, if one disregards the rather poor "translation" of the chemical terms, is superior to that dealing with the theory. A chemist could repeat the experiments described in the paper by following scrupulously the procedures which are quite specific and intelligible.

In summary, the reviewer believes that any good chemist, familiar with the subject matter and with a few peculiarities of the Russian language ("s" is translated "with," and "by" is translated "of") would find the translation usable.

In his formal report to me, Dr. Weiss did not comment specifically on the economics of the translation. He informed me, however, that he estimated that it took him four times as long to attain a complete understanding of the significance of the paper as he would normally expect to require to read and understand an average manmade translation.

Please feel free to call on us for any clarification of this report.

Sincerely yours,

JOHN SHERROD,
Chief, Science and Technology Division.

HUMAN TRANSLATION (4)

INVESTIGATIONS IN THE FIELD OF QUINOLINE AND ITS DERIVATIVES

XXI CONJUNCT CONDENSATION OF ARYLAMINES WITH HYDRACRYLALDEHYDE

B. I. Ardashev and V. I. Minkin

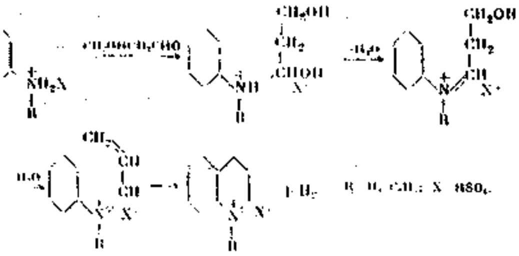
Novocherkassk Polytechnical Institute

It has previously been pointed out that hydracrylaldehyde (β-hydroxypropionaldehyde) can participate together with acrolein in the Skraup reaction [1]. Indeed, it is formed in significant amounts both during the decomposition of glycerin [2] and during the addition of water to acrolein at elevated temperature [3], the addition reaction being catalyzed by acid [4]. In connection with this, it was of interest to study the interaction of certain aromatic amines with hydracrylaldehyde with the aim of obtaining quinolines. With the exception of a patent to Chichibabin*[5], this reaction has not been described in the chemical literature.

We have established that under the usual conditions of the Skraup synthesis, namely, the dropwise addition of hydracrylaldehyde to the reaction mixture at 120-140°, quinoline is obtained in a yield of 15 %. If milder conditions are used, m-nitrobenzenesulfonic acid as the oxidizing agent and a dilute solution as the reaction medium, the yield of quinolines resulting from the use of hydracrylaldehyde reaches 50%. This permits the conclusion that when the Skraup reaction is carried out with glycerin in the usual manner, and also when acrolein is substituted for glycerin [6], one course of the reaction is interaction with hydracrylaldehyde formed from the glycerin or acrolein.

The reaction of hydracrylaldehyde with aromatic amines is now extended to diarylamines. Thus, from diphenylamine we were able to obtain a salt of N-phenylquinoline, and this reaction has been applied to other diarylamines. In this manner, it has been shown that it is possible to introduce secondary amines into a modification of the Skraup synthesis with hydracrylaldehyde.

Of the several possible courses of the reaction of hydracrylaldehyde with arylamines leading to the formation of quinolines, the most probable is that represented by a scheme in which the anil of hydracrylaldehyde is converted to the anil of acrolein, which then cyclizes to the quinoline



*The French spelling of the name is Tchitchibabine – Publisher's note.

Literature sources relating to the study of the mechanism of the Skraup reaction * and the experimental data obtained in the present work lead us to the conclusion that the Skraup reaction proceeds both through interaction with acrolein and through interaction with hydracrylaldehyde and, probably, glyceraldehyde **, and any of these courses can predominate depending on the starting products and on the conditions used

EXPERIMENTAL

The reactions were carried out in a three-necked flask fitted with a stirrer, a thermometer, and a reflux condenser in which was inserted a dropping funnel. The first experiments were carried out using hydracrylaldehyde itself, but later, in view of the fact that the results were identical, the diethyl acetal of the ethyl ether of hydracrylaldehyde was used, since it was more readily available [5]. In order to avoid excessive tarring and decreased yields, the aldehyde components were always twice-rectified, and the product boiling at 91-92° at 32 mm was used.

Experiment 1. Quinoline. 9.3 g of aniline was added to 25 ml of H_2SO_4 (d. 1.84) and 12.5 g of nitrobenzene. The mixture was heated to 120°, and 17 g of the diethyl acetal of β-ethoxypropionaldehyde was added dropwise over the course of 0.5 hour; during the addition, the temperature did not rise above 130°. After this, the mixture was refluxed for an additional 1-2 hours at 138-148° The quinoline was separated with potassium ferrocyanide [8]. The yield was 2.05 g (15.6%). B.p 234-240°; m.p. of picrate, 201°

Experiment 2. Quinoline. A solution of m-nitrobenzenesulfonic acid was prepared by Utermohlen's method [9] from 12.5 g of nitrobenzene, 50 g of 20% oleum, and 20 ml of water. 9.3 g of aniline was added to the solution; then, at 125-135° and over the course of 30-45 minutes, 20 g of the diethyl acetal of β-ethoxypropionaldehyde was added dropwise, after which the mixture was heated at 135-145° for another 1-1.5 hours. The yield of quinoline was 6.95 g (53.8%). B.p 233-238°; m.p. of picrate, 200°

Experiment 3. 6-Methoxyquinoline. This experiment was carried out similarly to Experiment 2. 6.74 g (43.4%) of 6-methoxyquinoline was obtained from 12.3 g of p-anisidine. B.p. 278-284°; m.p. of picrate, 217°.

Experiment 4. N-Phenylquinolinium perchlorate. 10 g of the diethyl acetal of β-ethoxypropionaldehyde was added, at a temperature of 120-130° and over the course of 15 minutes, to a mixture of 8.5 g of diphenylamine, 45 g of nitrobenzene, and 10 ml of H_2SO_4, and the mixture was heated for another 2 hours at 125-135° The product was separated in a manner similar to that used for the isolation of N-arylquinaldinium salts [10]. The yield of perchlorate was 1.72 g (11.2%). Recrystallization from water gave a product with an m.p. of 157°

Found %: C 58.75; H 3.79; Cl 11.83; ClO_4 33.78. $C_{15}H_{12}NClO_4$. Calculated %: C 58.93; H 3.95; Cl 11.60; ClO_4 32.69

SUMMARY

1. Quinoline derivatives were obtained by the reaction of hydracrylaldehyde with anilines.

2. It was shown that it is possible to introduce diarylamines into the Skraup reaction. N-Phenylquinolinium perchlorate was synthesized.

3. The mechanism of the reaction was discussed, and a scheme was proposed for the reaction.

LITERATURE CITED

[1] B. I. Ardashev, J. Gen. Chem., 16, 47 (1946).

[2] R. Oshil'd and P. Pi'l, Chemistry and Chemical Technology, 8 (91), 40 (1957)

[3] A. Nef, Lieb. Ann., 335, 219 (1905)

*For example, Iale pointed out that when acrolein is similarly used, only traces of quinoline are formed [7].
**We are presently investigating this course of the reaction

[4] F Piotert,* Organic Syntheses. Foreign Lit. Press, 3, 88 (1952).

[5] A E. Tchitchibabine. Fr. Patent 961, 466, May 28, 1951; Ch A, 49, 383 (1955).

[6] B. I Ardashev, Prog. Chem., 23, 46 (1954).

[7] H L. Jale, and J. Bernstein, J Am. Chem. Soc., 10, 254 (1948).

[8] N. S. Kozlov, J Gen Chem., 7, 1864 (1937).

[9] W P Utermohlen, J. org. Ch., 8, 544 (1943).

[10] V I Minkin and B I. Ardashev, J. Gen Chem., 28, 2556 (1958).**

Received November 21, 1957

*Transliteration of Russian – Publisher's note.
**Original Russian pagination. See C. B. translation