B A L L I S T I C   R E S E A R C H   L A B O R A T O R I E S

MEMORANDUM REPORT NO. 1147

MAY 1958

SURVEY OF THE FIELD OF MECHANICAL
TRANSLATION OF LANGUAGES

George W. Reitwiesner

Martin H. Weik

Requests for additional copies of this report will be made direct to ASTIA.

A B E R D E E N   P R O V I N G   G R O U N D ,   M A R Y L A N D

BALLISTIC   RESEARCH   LABORATORIES

MEMORANDUM REPORT NO.   1147

SURVEY OF THE FIELD OF MECHANICAL
TRANSLATION OF LANGUAGES

ABSTRACT

The report describes the present state of the art and presents
an analysis of the subject of mechanical translation of foreign
languages by means of high speed electronic computers. The results
of a survey of the field of mechanical translation and a bibliography
are included.

TABLE OF CONTENTS

I INTRODUCTION

An announcement was made in 1954 of the results of a Georgetown University-International Business Machine Corporation "experiment" in the translation of Russian into English by a digital calculating machine. Since then, several other organizations have become interested in the mechanical translation of languages, i.e. in the translation of text from one language (source) to another (target) by electronic digital data-processing devices, predominantly in the translation of Russian scientific written text into English.

In December 1957 the Computing Laboratory of the Ballistic Research Laboratories was requested by the Office of Ordnance Research to conduct a survey of the state of the art of mechanical translation. The results of that survey are included in this report.

The survey was conducted in two ways: by studying literature in the field and by interviewing workers in the field. All organizations known to be active in the field were asked to supply literature describing their efforts. Seven leading organizations in the field were visited for interview with the personnel engaged in the work; the seven were chosen from a dozen identified by the National Science Foundation in January 1958; the basis of choice being largely that of convenience in visiting the largest possible number within a reasonably limited itinerary.

II THE PRESENT STATE OF THE ART OF MECHANICAL TRANSLATION

There are two major aspects of the subject of mechanical translation
of foreign languages by machine methods. One of these pertains to the
"language" e.g. grammatical rules, syntactical analysis, dictionaries,
flow charts and programming. These considerations have been grouped
under the term "Methodology". The other major aspect pertains to machines,
i.e. character sensing equipment, information storage devices, printing
devices, and arithmetic or processing organs. The considerations per-
taining to this aspect of machine translation have been grouped under the
tern "Equipment".

A. Methodology

Several steps are required for the mechanical translation of a
text from one language to another: (a) pre-editing, (b) reading,
(c) analysis, (d) substitution, (e) writing, and (f) post-editing. Analysis
implies grammatical and syntactical analysis; substitution implies the
replacement of a word or group of words in one language with its or their
equivalent in another, through some form of dictionary look-up. Pre-
editing and post-editing imply the removal and insertion of distinctive
forms of expression in the languages concerned, so that the remainder of
the translation process can be routinized to a satisfactory degree. The
several steps need not be performed in time sequence in the order given
above. In a suitably organized translating process either pre-editing or
post-editing may be performed as a part of analysis. The degree of re-
finement of the finished product varies according to the extensiveness of
the pre-editing, analysis, substitution, and post-editing steps. In the
crudest case, the translation process consists of simply reading, substi-
tution, and writing. In the efforts of the several workers in the field,
varying intensities of emphasis are placed upon pre-editing, analysis,
substitution and post-editing. Effort is being devoted toward removal of
the pre-editing and post-editing steps. Many workers in the field have
agreed that pre- and post-editing can be eliminated when an operating
system has been developed.

A digital data processor, employed for language translation, is viewed as consisting of four components: input, storage, processor, and output. Output appears to offer no major problem for mechanical translation. A variety of types of printing devices exist. Storage capacity about equivalent to one million words of 40 binary digits each appears to be required to house a dictionary for any one field of science. This capacity is realizable in the magnetic tape systems of existing digital machines, but at discouragingly slow speed. Higher speed devices of sufficient capacity appear in the offing. The processor needs little conventional arithmetical power, but it must have considerable power to split apart and to combine characters and groups of characters and to effect comparisons among characters and groups of characters. Certain peculiar operations are undoubtedly called for. Currently available machines may prove inadequate either in types of operations available or in speed of performance, but there is yet no conclusive evidence that such will be the case. The input consists of sensing the characters of the source language, which have been manually keyed onto punched cards or tape. Mechanical readers for conventional textual formats are under development. Additional detailed information concerning equipment requirements for accomplishing mechanical translation is given in paragraph B "Equipment".

1. Summary of the Results of a Survey

From all indications, none of the organizations active in the field expects to have the capacity to translate an arbitrarily selected text in any one field of science for at least 6 to 12 months from the date of this report; several are hoping to be able to perform at least crude translations within that time; some do not expect to have working programs before at least a few years. Thus, at least a year should pass before an evaluation can be made of the worth of machine translations, however crudely or smoothly prepared, under currently contemplated procedures. No criterion has been established for such an evaluation, other than that which exists by implication: customer satisfaction.

The following paragraphs summarize the activity in mechanical translation in various geographical areas.

a. Boston Area

Both Harvard and MIT have conservative attitudes. A Russian to English dictionary is being constructed as the first step in the Harvard program; the grammatical analysis of German was the primary initial effort of MIT. Both universities anticipate relatively slow progress; both have good machine facilities in the form of powerful scientific computers.

b. Washington Area

Georgetown University workers are predominantly linguists. They are seeking the support of experienced programmers. Machine facilities abound in the Washington D. C. area. Georgetown has begun programming for an NBS-DOFL computer.

c. Seattle Area

Both linguists and engineers are active at the University of Washington; machine facilities, however, are limited. A study of over 100 texts in over 30 fields of science has yielded certain data for translation programs, e.g. data on multiplicity of meanings, storage capacity, etc.

d. Los Angeles Area

Ramo-Wooldridge Corporation and Rand Corporation are essentially systems analysts, receiving linguistic support from U. C. L. A. Both organizations have excellent machine facilities.

e. Other Domestic

Accounts of activity of twenty-two organizations that have performed some study in mechanical translation, are given below.

f. Foreign

Known foreign efforts in the field of MT are limited to two nations: The United Kingdom and USSR. Accounts of the efforts of three British organizations and of the Academy of Sciences in Moscow are given below.

2. Machine Translation Activity at Specific Domestic Agencies

a. U. S. Government Agencies

The following U. S. Government agencies have directly contributed toward progress in the field *of* machine translation:

        (1) National Bureau of Standards
             U. S. Department of Commerce
             Applied Mathematics Laboratory
             Washington 25, D. C.

The Applied Mathematics Division, Computation Laboratory, is planning to undertake a project to program translation of Russian into English using the DOFL-NBS IBM 704 computer. The work is to be supported by a grant from the U. S. Army Office of Ordnance Research for the first six months. The program is to be headed by Mrs. Ida Rhodes, who has a knowledge of Russian and is an expert programmer.

        (2) National Science Foundation
             1951 Constitution Avenue
             Washington 25, D. C.

The National Science Foundation published the first edition of a semi-annual report on "Current Research and Development in Scientific Documentation" in July 1957. The descriptions of work in progress contained in the issue was divided into:

1. Organization of Information

2. Equipment for Storage and Retrieval

3. Special Studies

4. Mechanical Translation

The contributing organizations in the field of mechanical translation were

        Birkbeck College, London

        California Institute of Technology

        Cambridge Language Research Unit, Cambridge, England

        Georgetown University

        Computation Laboratory of Harvard University

        International Telemeter Corporation

        Massachusetts Institute of Technology (2)

Ramo-Wooldridge Corporation

Rand Corporation

University of Michigan

University of Washington

The National Science Foundation sponsored a series of conferences to co-ordinate data processing within the government, including the armed forces. Much attention at these conferences was paid to MT. One conference was held on 10 February 1958. Dr. Adkinson, of NSF, was charged with the responsibility of keeping the Killian Committee (advisory to the President) informed on matters pertaining to data processing within the government.

     (3) U. S. Air Force
       Air Research and Development Command
       Rome Air Development Center
       Griffiss Air Force Base
       Rome, New York

       The U. S. Air Force, Air Research and Development Command, Rome Air Development Center technically supervised the Air Force sponsored contract for the construction of equipment designed specifically to accomplish machine translation of languages. Reports on this equipment are issued under Contract AF30(602)-1566 with International Telemeter Corporation and may be ordered from ASTIA by the following ASTIA AD-numbers:

97818-B	148541
97947	131384
114238	148542
114354	148543
114356	148544
114384	148545
114437	148546
131380	148547
131381	148548
131382	131383

The University of Washington (Subcontractor) is issuing a final report, and it is expected that it will be available for distribution at about the same time as this report (BRLM l147).

(4) U. S. Army
                        Pentagon
                        Washington 25, D. C.

          The Department of the Army supported work at the
University of Michigan under DA Contract DA-36-039-sc-52654. See subheading
"University of Michigan" for additional information (U. S. Army Signal
Corps).

              The U. S. Army Office of Ordnance Research is supporting
the National Bureau of Standards in an effort to obtain a useful trans-
lation of Russian scientific literature by machine methods. The joint
program includes (1) Formulation of grammatical rules (2) establishment of
an automatic dictionary (3) coding and programming and (4) machine solution
on the DOFL-NBS IBM 704 computer. The NBS team is headed by Mrs. Ida Rhodes.
(See Chapter II, Section A2, Subheading "National Bureau of Standards".)
(U. S. Army Office of Ordnance Research)

            b. Commercial Organizations

          The following commercial organizations have directly
contributed toward progress in the field of mechanical translation.

                    (1) Bell Telephone Laboratories
                        Murray Hill, New Jersey

          In 1954, K. H. Davis worked on machine recognition
of phonetic elements by utilizing a limited-vocabulary automatic word
recognizer for use in telephone activity. Oral language translation is a
possible application of the results of Davis' investigations.

            The Laboratories have also been engaged in character
recognition for automatic reading of telephone bills, routing slips and.
charges. (It is to be assumed that efficient reading of a printed page is
a requirement for rapid mechanical translation of foreign printed scien-
tific text.)

(2)   Haskins Laboratories
              305 E. 43 Street
              New York 17, New York

        During 1954, Mr. Franklin S. Cooper, of Haskins
Laboratories, Inc. worked on the problems of analysis and synthesis of
speech sounds and their perception. Consideration was given to possible
mechanical translation of the spoken work into other languages.

        (3)   International Business Machines Corporation
              590 Madison Avenue
              New York 22, New York

        During 1954, Peter Sheridan, of IBM, worked on the
use of the IBM 701 Computer for Russian - English conversion. This was a
joint effort by IBM and the Institute of Languages and Linguistics of the
School of Foreign Service of Georgetown University. Active persons included
Dr. Leon Dostert and Dr. Paul Garvin of the Georgetown staff. This joint
IBM-Georgetown effort is commonly referred to as the "Georgetown Experiment."

        Mr. Peter Sheridan presented a detailed account of
the computer programming for the IBM-Georgetown University experiment at
the 24 June 1954 session of the Association for Computing Machines conference
at Ann Arbor, Michigan. (See subheading "Georgetown University").

        (4)   The Rand Corporation
              Santa Monica, California

        During 1957, a group headed by David G. Hays worked
on the development of a body of Russian text on punched cards, with standard
procedures for editing, key-punching, grammatical coding, and the assigning
of English equivalents.

        A glossary has been prepared, in cooperation with
Andreas Koutsoudas of the University of Michigan. The glossary contains
all the inflected forms to be found in a specific text. Studies were made
of frequencies of certain word associations, sentence structure, and con-
text of prepositions and conjunctions. The glossary and text is to be
gradually increased. The work includes a study of the statistics of word
occurrences. Mr. Hays has built a 12,000 word glossary from 100,000 words
of running text in physics. Plans are to expand the vocabulary to 24,000
words in approximately a year. The Rand Corporation's approach is to obtain

                              14

a word-for-word translation first, then to expand this to include groups of words, then add the grammatical refinements and additional programming to produce "neat" translations. The Johnniac is being programmed for performing the translation. An IBM-704 with a 32,768-word store is available.

<div align="center">

(5) Ramo-Wooldridge Corporation
     Data Processing Systems Department
     Information Systems Division
     Los Angeles, California
</div>

Work accomplished by the Ramo-Wooldridge Corporation has been primarily in the field of translation of Russian to English. Much of this work was done under an Air Force Contract. ("Appendix A, Machine Translation" (unclassified), part of Ramo-Wooldridge's report to the Air Force on work under this contract, was submitted June 1957. Qualified requestors may obtain copies of the Appendix from ASTIA Document Service Center, Arlington, Virginia). The Ramo-Wooldridge Corporation has recognized that there are two main problems which must be solved: the linguistic problem inherent in the translation process; and the problem of mechanizing the translation process.

Early in the program at Ramo-Wooldridge, the possibilities of word-for-word machine translation were investigated. An experiment was conducted in which simulated word-for-word machine translation was accomplished. Redundancy and a high degree of ambiguity were two serious defects in the translated products. To overcome the redundancy and ambiguity, an intensive study of the fundamental linguistics involved in translation was undertaken. Means were needed for eliminating three types of ambiguity: syntactic, semantic, and idiomatic. Extensive empirical study gave sufficient understanding of these problems to enable the establishment of machine techniques for their solution, at least in part. It was found that ambiguity could be overcome by analyzing ambiguous words within their immediate context. The result would be "phrase-for-phrase" or "sentence-for-sentence" translation.

A special coding scheme was devised whereby each word contained in the machine dictionary is identified according to its various characteristics, i.e., part of speech, number, grammatical case, etc. Each word in the dictionary thus has a set of special codes which define its characteristics. Concurrently, translation rules were worked out by which words could be extracted from the machine dictionary, examined in their context, and then the correct word or words selected. These rules were devised in terms of the logical operations of a computer.

Syntactic ambiguity was resolved by examining a word in the code framework of adjacent words. A set of translation rules, believed practically complete but subject to refinement, was compiled for resolving syntactic ambiguity. Semantic ambiguity was partially solved by operating within a restricted glossary related to some particular subject or field. Within many subject fields this is an adequate solution. However, a more definitive solution may come from a comprehensive empirical study of language in action, i.e., large-scale observation of occurrence of particular English equivalents in association with certain other words in sentences. Idiomatic ambiguity was resolved by simply listing idiomatic expressions, their appropriate coding, and corresponding translation.

After devising and refining the special coding procedures and translation rules discussed above, a carefully controlled experiment was conducted in which simulated machine translation on a phrase-for-phrase basis was performed. On the whole, the translation product was quite satisfactory, except for failure to resolve semantic ambiguity completely in all instances.

Concurrently an intensive systems study was made of the problem of mechanizing the translation process; i.e., of the computing facilities and related equipment required to accomplish machine translation. This systems study revealed that existing, available equipment could be assembled which would prove adequate. General purpose digital computers, magnetic tape, Flexowriters, and other available equipment may serve the purpose. However, this systems study also revealed that special purpose equipment, designed specifically for machine translation use, probably would provide machine translation at a lover cost.

At present writing, Mr. Swanson is working on a procedure whereby feedback from a grammatical analysis approach can be directed to a linguist for improvement of technique. Plans are to have an IBM 704 program running in perhaps three to six months. Mr. Swanson is interested in studying logical structure for a machine designed specifically for language translation.

(6) Telemeter-Magnetics, Incorporated
2245 Pontius Avenue
Los Angeles 64, California

During 1956, Gilbert W. King worked on stochastic methods of mechanical translation, dealing with the problems of syntax, grammar, and context sensing for selecting the proper translation from a group of possible translations by a probabilistic technique. Much emphasis was placed upon context and dealing with on-the spot analysis of an entire sentence.

Under a U. S. Air Force contract, (See Chapter II, Section Aa, Subheading "U. S. Air Force"), Dr. King developed a memory device to be used for a mechanized dictionary of Russian - English entries. In brief, the device consists of a photographic disc on which are recorded the coded dictionary entries. The disc has a capacity of 3×10^7 bits of information, access time of 50 milliseconds, and a reading rate of 10^6 bits per second. Dr. Reifler, of the University of Washington, is preparing the dictionary entries to be stored on the disc. A paper entitled "The USAF Automatic Language Translator Mark I", describing this equipment, was presented at the IRE National Convention in New York on 27 March 1958, by G. A. Shiner, Rome Air Development Center, Griffiss Air Force Base, Rome, N. Y. (See Chapter II, Section B, Paragraph lb(2) "USAF Mechanical Translator Mark I").

c. Educational Institutions

The following educational institutions have directly contributed toward progress in the field of machine translation:

(l) California Institute of Technology
Pasadena, California

During 1957, Mr. Toma worked out experimental programs for machine translation of Russian texts with the Datatron Computer. Pertinent dictionaries are transferred from tapes to high-speed memory. Logical sequences are handled as semantic units of 2 or 3 words. Mr. Toma is enlarging and rewriting the program for larger computers. He has recently joined the Georgetown University workers.

<div style="text-align:center">

(2) Georgetown University
Institute of Languages and Linguistics
1719 Massachusetts Ave. N. W.
Washington, D. C.

</div>

During 1954, much of the work, particularly the original "Georgetown - IBM Experiment", in which MT was accomplished on a limited scale on a IBM 701 computer, was supported by Georgetown University, the International Business Machines Corporation, and the Rockefeller Foundation.

By November, 1956, a group of linguists, under the direction of Dr. Leon Dostert, were assigned to execute the following program: (l) Gather experts and organize a series of problem-seminars (2) focus attention on Russian chemistry text (3) analyze subject material from lexical, morphological and syntactic viewpoint and (4) perform coding for machine operation and translation into English.

During 1957, most of the work in connection with the Project in Machine Translation at the Institute of Languages and Linguistics of Georgetown University was organized in the form of a series of Seminar Work Papers. Each of the papers covers a specific aspect of activity in MT of Russian to English. Although emphasis has been placed on the translation of Russian chemistry text into English, most of the work is applicable to Russian scientific literature in general. The Bibliography (Section III of this report) lists the work papers published in the field of MT by the Georgetown Group.

This program of research, which is supported in part by the National Science Foundation, is aimed at the development of codes for lexical data and the rules and sub-rules of operational syntax necessary for mechanical translation of complete texts from Russian into English. The research is focused on texts in the field of organic

chemistry, using the Soviet publication <u>Journal</u> <u>of</u> <u>General</u> <u>Chemistry</u> from 1952 onward as primary source material. About 5,000 - 6,000 simple and compound words are being processed in terms of meaning equivalence and systematized coding for machine manipulation.

Three approaches to MT are being pursued at Georgetown. The one which shows the most promise, or an approach utilizing the best portions of all three, will be adopted and machine solution will be attempted.

One approach is dubbed the "experimental approach," which proceeds empirically from the specific pattern to the general formulation. From a semi-analytical and semi-empirical approach, the staff members expect to formulate a series of generalizations which can be turned into machine instructions. Another group has developed a machine technique based on grammatical analysis of Russian, to resolve what may be called "internal structural ambiguity." This is the "code-matching" technique. A third hypothesis seems to be purely an empirical approach, i.e. sentence by sentence. The argument here is that exhaustive analysis of a series of sentences in a continuous text will yield general rules.

At the present writing, Professor Zarechnak is engaged in an extensive study in the fundamentals of grammar. Professor Garvin has prepared detailed preliminary flow-charts for grammatical analysis accommodating Russian - English translation and is planning the programming of his work on the National Bureau of Standards' IBM 704 computer. Mr. Austin and Miss Lukjanow are developing a matching technique.

(3) Harvard University
     Computation Laboratory
     Cambridge, Massachusetts

The machine translation project group at Harvard, headed by Dr. Anthony Oettinger, is engaged in the preparation of a complete mechanical dictionary, to be placed on a machine in order to obtain a rough translation. The reader could then decide whether a more accurate translation is desirable. Arrangements are made to permit

enlargement of the dictionary and extend the simpler substitution task to include grammatical rules in both the source and target languages. With this approach, it is expected that a practical translation can be obtained, which then could be improved over the years as grammatical analysis is further developed and the dictionary expanded.

At the present writing, Dr. Oettinger, Computation Laboratory Staff, is planning for the trial operation of an automatic Russian - English dictionary with Harvard's Univac I Computer.

(4) Indiana University
     Bloomington, Indiana

D. G. Ellison worked on a mathematical model, which, if executed by a computer, would cause the computer to "learn" problems of the type represented in MT. He also worked on a computer which could read printed characters and correct some misprints when first encountered. The computer's "learning" of repeated misprints of the same type would gradually assume significance and become "intelligence".

(5) Massachusetts Institute of Technology
     Department of Modern Languages,
     and Research Laboratory of Electronics
     Cambridge 39, Massachusetts

Serious effort in the field of mechanical translation started in 1951 when Yehoshua Bar-Hillel at MIT became the first full-time worker in the field.

The First International Conference in MT was held at MIT in 1952. Agreement was reached at that conference that word-for-word translations were inadequate and syntactic problems would have to be handled by machine.

During 1953, work at MIT was supported jointly by the Massachusetts Institute of Technology, the U. S. Army Signal Corps, the Air Material Command, the Office of Naval Research, and the Rockefeller Foundation.

In 1954, Morris Halle and William N. Locke were working on automatic identification of speech sounds and possible use of such identification as input to a mechanical translator.

In 1954 and 1955, Victor H. Yngve worked on a scheme for applying statistical techniques to provide a description of syntax, on syntax and multiple meaning problems in general, on German - English translation, the application of information theory of problems of MT, and lexical research on German to English for partial translations. Attention was also devoted to a study of methods for securing wide publication of translations, and securing answers to financial and copyright questions.

By 1955, V. Yngve concluded that the direction of effort in MT should be to (l) reduce the size of the dictionary by formulating rules of grammar. (He contends that rules of grammar are quite stable) (2) store complete sentences when necessary, in order to avoid ambiguity when translating idioms, etc. (3) place all possible constraints on expression so as to reduce the number of possible sentence translations.

By November 1956, plans were completed to develop an adequate description of the German and English languages in order to develop proper word order in a German translation of an English sentence by machine methods.

N. Chomsky worked on a theory of grammar that gives many new and powerful insights into the structure of language.

J. Applegate worked on the detailed structure of the German verb phrase.

R. Lees and G. H. Matthews have looked into the structure of the German verb phrase.

During 1957, under a grant from the National Science Foundation, V. Yngve carried out basic research on sentence structure, with a view toward mechanical translation of languages. He concentrated on the grammar and syntax of both German and English languages.

Lexical research is presently under way to complement the syntactic research. Much of the time of the research team headed by V. Yngve has been spent on a description of German and English for mechanical translation. A number of specific topics have been studied. Among these are sentence structures, classification of nouns, noun modifiers and verbs, participial phrases, adverbs, and co-ordinating conjunctions and other grammatical structures. Investigation is being made of correspondence in sentence structure between German and English.

V. Yngve's present approach to the MT problem is to place major emphasis on grammatical analysis and leave the problem of preparing a complete dictionary for a later date.

The group at MIT have been publishing a periodical called "Mechanical Translation", to serve as an organ for reporting progress and publishing articles in the field of MT.

(6) State College of Washington
     Pullman, Washington

As early as 1954, Richard H. Lawson, Department of Foreign Languages became interested in the frequency of idiomatic expressions in scientific German. Much of his interest centered around idioms which are not translatable by the microglossary techniques utilized in mechanical translation.

(7) University of California at Los Angeles
     Los Angeles, California

Much of the early work at U. C. L. A. performed by William E. Bull, Charles Africa and Daniel Teichroew was supported by the Office of Naval Research and the Rockefeller Foundation.

Later, Messrs. William E. Bull, Charles Africa, and Daniel Teichroew completed a linguistic analysis of language, paying particular attention to the "word" as an identifiable entity, and its significance in MT. William Bull prepared word counts on some 200 titles over a period of several years.

Also in 1954, Kenneth E. Harper of the Slavic Department, became deeply interested in a study of Russian vocabulary and syntax as applied to MT.

By November 1956, K. Harper performed contextual analysis in word-for-word mechanical translation. Experiments with word-for-word mechanical translation of Russian scientific literature proved this to be an unsatisfactory approach. Most emphasis was given to a study of nouns. Efforts were made to establish an ideoglossary based on actual rather than potential behavior of words, since in most cases, only one technical meaning is usually assigned by technical writers to any given word.

At present writing, he is working on several approaches to MT. Professor Harper is serving as a consultant to Ramo-Wooldridge Corporation and the Rand Corporation on the matter of Russian - English machine translation.

(8) University of Chicago
Chicago 37, Illinois

W. H. Meyer and J. Wilkinson have gathered an extensive collection of papers dealing with the general theme of language and meaning. Particular emphasis is on "structural" theories of meaning and the revision of theoretical grammar in light of structural meaning and the concepts of modern logic. Much of this work has been done with a view toward application in MT.

(9) University of Michigan
Engineering Research Institute
Willow Run Laboratories
Ypsilanti, Michigan

In 1955, the University of Michigan initiated research in the analysis of language structure for mechanical translation. Andreas Koutsoudas and R. Korfhage began work on Russian text selected from the Journal of Experimental and Theoretical Physics. Most of the effort was devoted to machine handling of multiple meanings.

Mr. Koutsoudas is fundamentally interested in research in linguistics.

Much of the work in MT at the Willow Run Laboratories, University of Michigan, was supported by DA Contract No. DA-36-039-sc-52654. A set of 15 rules were established and tested to determine their adequacy to handle the multiple meaning problem. The approach has been successful. In essence, multiple meanings are listed in sequence. When a word with multiple meanings is encountered, translation is postponed until the local structure is analyzed, then the proper meaning is selected. Sometimes the same meaning has to be stored more than once. Some preliminary tests were made by persons who had no knowledge of Russian. They were given the set of rules, a short dictionary and asked to proceed with a translation of a sample text.

During 1956, A. Koutsoudas and R. Machol worked on the syntactic and grammatical aspects of translating Russian into English. At that time 128 pages, consisting of 64,000 running words, from the Journal of Experimental and Theoretical Physics were coded into punched cards. Experiments were carried on in which technicians simulated a computer in translating according to the stated rules. Answers were being sought to such questions as: How many different words will be found in a million running words? How many new words will be found in a second sample equally large? How *many* words must there be in a dictionary to ensure having 99% of the words in a sample randomly chosen from a certain field? It was recommended that a suitable standard type font be chosen by publishers of technical journals in order to facilitate mechanical translation.

A.Koutsoudas and R.Korfhage came to the conclusion that the present concept of conventional grammatical categories, i.e. verb, noun, adjective, etc. will be completely abandoned.

During 1957, A. Koutsoudas directed his work toward the design of a process to translate foreign text accurately into precise, stylistically correct English without the necessity of human pre- and post-editing. An empirical approach was used to solve linguistic problems in

certain areas of difficulty. Particular emphasis was placed on the problem of multiple meaning. Attention is now being devoted to the preparation of special dictionaries with as many as ten English equivalents, the investigation of Russian grammar, and the conduct of word frequency studies.

(10) University of Pennsylvania
      Philadelphia 4, Pennsylvania

At present writing, Mr. Zelig Harris is working on transformational analysis and canonical forms for mechanical translation. He is investigating the feasibility of mechanizing the application of a transformational analysis technique to scientific texts in English, with the thought that such a mechanized procedure would represent a useful step in an information retrieval system. This work may have applications in the field of mechanical translation.

(11) University of Washington
      Department of Far Eastern and
      Slavic Languages and Literature
      University of Washington
      Seattle, Washington

In April 1951, Professor Erwin Reifler prepared a reply to Dr. Warren Weaver's original suggestion that mechanical translation was possible through the use of high speed electronic digital computers. The reply entitled, "Studies in Mechanical Translation No. 1", described the various theoretical and practical aspects and problems which one must consider when attempting to translate languages by machine methods. Some of these considerations are pre-editing, labelling, coding, idioms, word order, grammatical meaning and mechanization of dictionaries. Here, Professor Reifler cites the "Golden Rule" of MT; namely, "The target language need not be more intelligible than the source language".

In 1952, Professor Reifler completed an analysis of German language for MT purposes. Proposing that there is a requirement for pre-editing, Professor Reifler concluded that a pre-editor is not required for German to English Translation. "With the confirmation of Drs. Bar-Hillel, Bull and Oswald, the German pre-editor faded completely into oblivion".

25

In June 1952, Professor Reifler presented an analysis of the MT problem, including post- and pre-editing. Many translation situation problems were analyzed in his paper presented at the MIT Conference on Mechanical Translation. Several proposals were made for solving many problems, such as the problem of translation when several meanings are involved. A suggestion to code source language parts of speech by capitalization of successive letters of words, like the German technique to capitalize all nouns. The second letter capitalized could indicate verbs, etc. In a second paper presented at the conference, Professor Reifler approached the problem of MT by means of grammatical analysis, touching on Dodd's suggestion for the adoption of "model English", using forms like "bringed" for "brought" and "I be, you be, he be, etc". Every form of the past tense is preceded by "did" and the future form is preceded by "will".

Continuing his work in MT, Professor Reifler presented a paper on "MT, Its Psychological Aspects and Its Significance for the Human Society" before the Second Conference on General Semantics, St. Louis, June 12, 1954.

In January 1954, Professor Reifler, in his "Short Report on the Present Stage in the Development of Mechanical Translation" gave a brief history of early interest in MT, described the first translation machine, generally described the MT problem, presented German syntactic problems related to MT and discussed the requirement for a pre-editor, and pointed out that ultimately all translation can be mechanized without human intervention at any stage. He continued his work in the treatment of words of dual nationality, the mechanical determination of proper names, the mechanical determination of non-grammatical meaning, and a further refinement of the MT form-class filtering system. He performed studies in linguistic analysis and comparative semantics with a view toward application to MT. He attacked the problem of meaning, indicating that meaning must be taken into consideration when linguistic analyses are made. He completed an analyses of form classes (Lexical forms that have any

function in common belong to a common form class). Many presently used grammatical terms, such as adverb, verb, adjective, preposition, etc. should be discarded in favor of "function" words, and new "MT form classes" should be adopted. Studies were made of "mutual pin-pointers" and engineering expediency. The conclusion is that MT form classes and linguistic form classes (i.e. parts of speech etc.) differ. The development of form classes for MT is somewhat governed by engineering considerations. MT classes are more inclusive not only because they are based on a consideration of the total meaning of linguistic forms but also more realistic in view of the fundamental semantic and formal nature of every linguistic sign.

During this period, Robert E. Wall studied the engineering aspects of the machine translation of language. The translation process was divided into four steps: encoding, memory search, logical operations, and decoding. State of Technology indicates that solutions are obtainable for performing each of these steps.

Working on the Mechanical Translation Pilot Model Project, Professor Reifler produced a useful approach to the Translation of German into English. Source language modification and machine technology were primary considerations. Dr. Thomas M. Stout of the E. E. Department, planned an MT Pilot Model, designed to demonstrate the engineering feasibility of Professor Reifler's approach to MT. A treatise by Professor Einstein, "On the Special and General Theory of Relativity" was used as a sample text. A set of rules, a code and a dictionary were proposed to use on the machine. An intelligible output was obtained, i.e. intelligible to a person familiar with the subject matter.

In 1956, Professor W. Ryland Hill, E. E. Department, Dr. Lew R. Micklesen, Professor of Russian, analyzed the ability of an International Telemeter Corporation specially designed translation machine to translate Russian into English. Problems were solved in the translation **of** scientific Russian in the field of idioms, syntax, and words with multiple meanings. At that time, it was agreed that the stage in machine

technology had been reached where a machine, specifically designed for translating, can be constructed. The basis for design was the photoscopic disc, capable of storing 30 x 10^6 black and white areas in concentric tracks on a transparent disc.

Dr. Micklesen worked out a complete tabulation of all subclasses of Russian paradigmatic form classes and determined the number of distinctive forms in each paradigmatic set. He investigated the process of compounding in the Russian language and prepared proposals for the economical dissection of compounds by machine, and he made an exhaustive analysis of MT form classes of the Russian language, the pre-requisite for the mechanical determination of intended grammatical and non-grammatical meaning.

In April 1957, Mr. Aristotelis D. Stathacopoulos demonstrated the feasibility of performing noun phrase analysis on the IBM 650 computer. A specially prepared program makes use of logical operations to identify and process noun phrases in English Text. A similar approach could be used for the grammatical analysis of source language for subsequent translation into a target language.

By July 1957, under Air Force contract, Professor Reifler, Dr. Micklesen and Professor Hill prepared the information which would have to be placed in a machine memory for automatic translation from scientific Russian texts into English. Work on the project was di-vided into two phases:

1. The Initial Project, financed by a grant of $30,000, began on 16 June 1956 and was completed by 15 March 1957. It dealt with the linguistic investigation and analysis of 111 Russian texts from 31 fields of science, and supplied approximately 13,000 Russian - English entries consisting of Russian "semantic units" belonging to the technical and general language vocabulary occurring in these texts, and additional lexical units selected from high frequency lists, and their target equivalents.

2. The Expanded Project, financed by an additional, much larger grant
   began on 16 March and was completed by 15 October 1957. Its ob-
   ject was the supplementation of all paradigmatic forms to the
   13,000 Russian "semantic units" and the supply of their target
   equivalents. This will increase the number of entries to about
   160,000.

This material was prepared for the first trans-
lation machine which does not have the logical equipment necessary for the
automatic resolution of grammatical and non-grammatical problems arising
from the discrepancies between the Russian and the English Language. (The
machine was built by the International Telemeter Corporation supported by
the USAF). Consequently, devices for the automatic reshuffling of word
order and the automatic pinpointing of intended grammatical and non-gram-
matical meaning are not yet included. Russian word order will be retained
in its output, which also will be cluttered up with "strings" of target
alternatives. But the study of predictions of the expected output has
revealed many opportunities for considerable output improvements by changes
in the original lexicography which take advantage of the practically un-
limited storage capacity of the memory device. Furthermore, the engi-
neering members of the project, in cooperation with the linguistic members,
worked on the elaboration of machine programs in preparation for the con-
struction of additional logical equipment to be included in an improved
translation machine. Professor Reifler has also prepared in improved index
to a Chinese dictionary, so that its contents can be more readily available.

Professor Reifler and his staff feel that about a
4×10^7 bit storage unit should suffice for both dictionary and programming
for translation in any one field of science. They feel that multiple
meanings generally are limited to three. They propose to translate idioms
as semantic units. They propose to translate only monoglots and to trans-
literate polyglots (e.g. Russian ΔOKTOP becomes English DOKTOR)

(12) Western Reserve University
                   Center *for* Documentation and Communication Research
                   School of Library Science
                   Cleveland, Ohio

          Dr. J. W. Perry, Director of the Center for Documen-
tation and Communication Research views mechanical translation as part of
a broader field. Dr. Perry's work has been in the field of information
organization, selection and retrieval by machine methods. Much that he
has accomplished has possible application in MT. For example, the Western
Reserve University, School of Library Science's Searching Selector has
direct application to MT, since automatic retrieval and automatic trans-
lation have common problems.

          The Western Reserve Searching Selector permits an
exceptionally wide range of concepts to be used in defining and conducting
searching operations. Thus, the scope of a search may be defined not only
in terms of specific substances, devices, attributes, processes, conditions,
organisms, persons, locations, etc., but also in terms of generic concepts
and their relationships to specific terms. Furthermore, observational
relationships, for example the roles in a given experiment or situation of
various substances, devices, etc, taken either specifically or generically,
may also be designated as points of reference in defining searches.

          This wide range of possibilities is accomplished
by the ability of the Western Reserve Searching Selector to detect combi-
nations of symbols and combinations of combinations at a multiplicity of
levels. At each level, combinations may be defined in terms of logical
product, logical sum, logical difference or derived complex logical
relationships. The different combinational levels may be thought of as
analogous to the combining of letters to construct sentences, sentences
to construct paragraphs, etc. The machine is able automatically to detect
the start and end of each organized symbolic unit analogous to word,
phase, sentence, or paragraph.

An information requirement is analyzed in terms of appropriate specific and generic terms, role indicators and logically defined relationships between them. The information requirement is thus analyzed on the same basis as is used to record the information contents of documents in the form of encoded abstracts. The searching step as performed by the Searching Selector consists of a series of logically defined matching operations involving the common set of terms used for analyzing the information requirement and the information content of documents. The Searching Selector has been designed so that ten searches may be conducted simultaneously. Such searches may be interrelated as to scope or they may be completely independent. A Flexowriter types out the serial number and bibliographic reference of the wanted abstract.

3. Machine Translation Activity at Specific Foreign Organizations

The following foreign organizations have directly contributed toward progress in the field of machine translation:

    a. Academy of Sciences
       Moscow, U. S. S. R.

The majority of work on machine translation performed in the Soviet Union has been done under the auspices of the Academy of Sciences in Moscow. The following remarks constitute a summary of the accomplishments and attitudes of the personnel at the Academy.

In November, 1956 D. Panov contended that both the lexical meaning and the morphological shape of the word can and should be used in analyzing the text. Work was based on (1) maximum separation of the dictionary from the translation program, (2) analysis of foreign languages and synthesis of Russian, (3) storing all dictionary words in basic form, from which Russian text may be synthesized, (4) storing all constant grammatical properties of words in the dictionary, (5) determine multiple meanings from context. Work began by making an analysis of Milne's "Numerical Solution of Differential Equations". Practical experiments were carried out on the BESM. The BESM is the most powerful known Russian electronic computer; its power approaches that of the IBM-704 and Univac Scientific 1103B computers. A dictionary of 952 English and 1,073 Russian words was compiled. Words are "found", in the machine by matching assigned numerical codes. Starting

with input English sentences, the entire translation process has been carried out automatically. Preliminary research work was carried on by I. K. Belskaya, philologist-in-chief, and by mathematicians I. S. Mukhin, L. N. Korol'ev, S. N. Razumovskii, G. P. Zelenkevich, and N. P. Trifonov. Research has been started to machine translate German, Chinese and Japanese into Russian. The input problem for Chinese and Japanese has been solved easily by using the Chinese Telegraph code. The work in German is being carried out by G. P. Zelankevich and E. A. Khodzinskaya under the direction of Belskaya. The Chinese effort is being made by A. Zvonov and V. A. Voronin and Japanese is being handled by M. B. Efimov.

The publication which best describes the advances in MT made by the Russians up to 1956 is the USSR Academy of Science's Report "Certain Problems of Automatic Translation" by I. K. Bel'skaya, L. N. Korolev, I. S. Muhkin, D. Yu Panov, and S. N. Razumovskii, Vol. 26, No. 12, 1956. This report belittles American efforts and considers the American approach to MT to be foolhardy. The report quotes Warren Weaver, out of context, as saying "A book written in Chinese language is a book written in English but coded in Chinese code, and in order to translate from one language to another we must descend from each language to a general basis of communication between people, to the existing but still undiscovered, universal language and then ascend to the other particular language". The USSR report then goes on to say that this method can hardly produce a solution.

Quite successful results have been reported in the translation of English into Russian on the BESM System, utilizing a numerical intercode, a regular automatic dictionary, a polysemantic-idiomatic dictionary, a set of instructions for analysis and synthesis of the source and target languages, and a numerical interlanguage using the binary look-up procedure by placing words in numerical order. The number of compares is $\log_2$ of the number of entries. For 30,000 entries, this amounts to about 15 audits. Finally, the Academy of Sciences plans to solve the problem of machine translation from one foreign language to another, using Russian as the "inter-language".

E. Z. Lyubimskiy, S. S. Kamynin and M. I. Filippova, of the Department of Applied Mathematics, Academy of Sciences, presented the principles of automatic reading of machine symbols at a conference in May 1957 in Moscow. A program has been developed for reading latin letters for the "Strela" machine. Automatic reading of the text considerably facilitated the feeding of the information into the machine and could also be utilized for automatic type setting.

At the same conference, L. S. Levinskiy devoted a paper to an automatic device which would convert printed text into sounds of speech.

Professor A. A. Lyapunov reported on the "General Problems of Machine Translation".

V. V. Ivanov spoke on the linguistic problems of creating a machine language for an information machine and suggested an approach to the creation of an abstract language by means of evolution of special "languages" for the individual fields of science, as well as by creating simplified conditional models of actual languages.

G. M. Vleduts, V. K. Finn, N. M. Yermolayev and Y. A. Shikhanovich reported on their work on information language on the fields of chemistry and geometry.

V. K. Finn and D. G. Lakhuti spoke on semantic requirements for an information language.

G. V. Chekova reported on experiments for translating mathematical texts from French to Russian using the "Strela" machine.

A decision was made by the Academy to publish a Soviet periodical similar to "Mechanical Translation". It was indicated that machine translation would become of great significance in Soviet science and economy.

T. N. Moloshnaya gave a description of preliminary investigations which are being carried out for developing rules for machine translation from English into Russian. The linguistic problem encompasses (1) orthography (2) morphology (3) lexicology (4) phraseology (5) syntax.

A great deal of work was devoted to establishing classes of words. In particular, 19 classes of words in English and 17 in Russian were determined. Methods were established in which an English sentence could be "rolled up" and a Russian sentence "unrolled" by coding the various word classes and accounting for word order. Dr. Panov's group consists of approximately 500 mathematicians, linguists, and clerical personnel, all working on machine translation of foreign languages into Russian and translation between foreign languages with Russian as an inter-language.

      b. Birkbeck College Research Laboratory
          University of London
          Department of Numerical Automation
          London W. C. 1, England

During 1954, A. D. Booth and R. H. Richens worked on the improvement of input-output mechanisms and prepared micro-glossaries for scientific topics for use in mechanical translation.

Later, Translations from French into English and German into English were made on the APEXC Computer, using micro-glossaries and stem-ending procedures. Much of the work was supported by the Nuffield Foundation.

During 1956, work was done by L. Brandwood of the Birkbeck College Computational Laboratory on the mechanical translation of French into English. A small dictionary was established and some syntactical instructions were stored in the machine memory. Only stems of conjugated and declined words were stored and the endings were added after textual analysis of the English. At this time the test had not yet been performed on the machine. It was expected that the results would show whether the various forms of words could be identified by machine and a suitable translation obtained.

      c. Cambridge Language Research Unit
          20 Millington Road
          Cambridge, England

In 1956, the Cambridge Language Research Unit became concerned with the analytic investigation of language and in particular with a correlative study of the descriptive-linguistic, logical, algebraic and other notational characteristics of natural languages and of translation between natural languages by machine methods.

In recent years, M. Masterman investigated the potentialities of a mechanical thesaurus, A. F. Parker-Rhodes investigated a linguistic basis of the thesaurus-type mechanical dictionary and its application to English preposition classification, and R. H. Richens investigated a program for mechanical translation between any two languages viz an algebraic interlingua."

Under a grant from the National Science Foundation investigations are in progress to determine the possibility of basing a mechanized syntactic program on algorithms rather than specific procedures for each word. Fields of interest include a mechanical thesaurus, combinatory logic, finite lattice theory, syntactic analysis, and construction of comprehensive mechanical dictionaries and programming schedules. Mrs. M. M. Braithwaite is Project Director.

       d. City and Guilds College
         South Kensington
         London S. W. 7, England

In 1954 Mr. Colin Cherry was working on problems connected with the establishment of a description of languages which would yield classifications better suited for mechanical translation than ordinary academic grammar.

B. Equipment

  1. Computing Systems

Experiments in language translation have been or will be performed on several general and special purpose electronic computing systems at a number of organizations in the U. S. and some abroad, particularly in the United Kingdom and the U. S. S. R. Translation in a crude and hardly usable form has been accomplished. Certain phases of mechanical translation programs have been conducted in connection with the following machines:

a. General Purpose Computers

    (1) APEXC

        Programs were run on the APEXC System at Birkbeck
College Research Laboratory for translation from French and German into
English. (See Chapter II, Section A2, Subheading "Birkbeck College
Research Laboratory"). The machine, <u>A</u>ll <u>P</u>urpose <u>E</u>lectronic <u>X</u>-ray <u>C</u>omputer,
is a medium speed computer with a pulse repetition rate of 50,000 Kc.,
with an add-substract time of 500 microseconds.. The word length is 32
binary digits. Storage is magnetic drum. The results obtainable from a
machine of such limited storage capacity and speed leaves much to be
desired.

    (2) BESM

        The BESM (Transliteration of Russian Initials for
Highspeed Electronic Computing Machine) is the most powerful of the Soviet
large scale high speed electronic computers. Several practical experiments
in mechanical translation were performed on the BESM at the Academy of
Sciences at Moscow. (See Chapter II, Section A2, Subheading "Academy of
Sciences"). According to S. A. Lebedev of the Academy of Sciences, the
BESM has the "faculty" of translating from one language into another. It's
memory device store words alphabetically and is capable of discriminating
among several different translations for a particular word. The word-for-
word translation thus obtained is edited by the machine with an eye to the
grammar and syntax of the language. The BESM has 5,000 tubes with 1,024-
words of high speed storage backed up by magnetic tape and high speed photo
print output.

    (3) DATATRON

        Programs for accomplishing machine translation of
Russian Text were prepared for the DATATRON system at The California
Institute of Technology (See Chapter II, Section A2, Subheading "California
Institute of Technology"). The DATATRON, manufactured by the Electrodata
Division of the Burroughs Corporation, medium scale, medium speed binary
coded decimal computer. The word length is 10 decimal digits and sign,

36

available in both floating and fixed point mode of operation. Add time is
169 microseconds, excluding access. Four 20-word "high speed" loops of
magnetic drum storage and 4,000-words per unit drum are available. This
is supplemented by Magnetic Tape and the "Datafile" systems. (A complete
description of the DATATRON is given in BRL Report 1010, "A Second Survey
of Domestic Electronic Digital Computing Systems", June 1957).

    (4) IBM 650

    In April of 1957, Mr. Aristotelis D. Stathacopoulas
prepared a program for performing noun phrase analysis of English text
on the IBM 650 Computer at the University of Washington. (See Chapter II,
Section A2, Subheading "University of Washington"). The IBM 650 medium
speed, 10 decimal digit per word computer with magnetic drum, magnetic tape
and magnetic disc (high capacity) storage. Sixty words of 96 microsecond
random access storage is available. The tape storage is compatible with
other IBM systems. Add time, excluding access time, is approximately 720
microseconds. (A complete description of the IBM 650 System is given in
BRL Report 1010 "A Second Survey of Domestic Electronic Digital Computing
Systems").

    (5) IBM 701

    An experiment, demonstrating the feasibility of per-
forming mechanical translation on the IBM 701 Computer, was performed by
an IBM-Georgetown University joint staff. (See Chapter II, Section A2,
Subheadings "Georgetown University" and "International Business Machines
Corporation"). (See BRL Report 1010 for complete description of this system).

    (6) IBM 704

    Professor Garvin of Georgetown University has pre-
pared a detailed preliminary flow chart for grammatical analysis accommo-
dating Russian - English Translation and is planning the programming of
his work on the National Bureau of Standards' IBM 704 Computer. (See
Chapter II, Section A2, Subheading "Georgetown University"). The IBM 704
Computer is a high speed, large scale computer with multiples of 4,096-
words of 12 microsecond random access magnetic core storage, with fixed

and floating point modes of operation. Magnetic tape storage, high speed input and output, and data conversion equipment are available. (See BRL Report 1010 "A Second Survey of Domestic Electronic Digital Computing Systems" for a complete description of this machine)

(7) JOHNNIAC

Programs are being prepared for performing mechanical translation on the JOHNNIAC Computer at the Rand Corporation. The JOHNNIAC is a fixed point, single address, parallel computer with a fixed word length of 40 binary digits and a 15 microsecond access storage of 4,096-words. A drum stores 12,288 words. Add time, excluding access, is 10 microseconds. (See Chapter II, Section A2 "The Rand Corporation"). (A complete description of the JOHNNIAC is given in BRL Report 1010).

(8) UNIVAC I

Dr. Oettinger plans to use the UNIVAC I Computer at Harvard for a trial operation of an automatic Russian-English Dictionary. (See Chapter II, Section A2, Subheading "Harvard University"), The UNIVAC I is a large scale, general purpose high speed serial electronic computer. Its word length is 11 decimal digits plus sign, operating in a fixed point mode. Add time is 282.6 microseconds. High speed storage is 1,000-words of acoustic mercury with an access time of 40.4 to 404 microseconds. The high speed storage is backed up by magnetic tape. High speed input and output (including high speed printer) and off line conversion equipments are available. (See BRL Report 1010 "A Second Survey of Domestic Electronic Digital Computing Systems" for a complete description of the UNIVAC I).

(9) STRELA

The Russian STRELA automatically senses characters of latin text. It is located at the Academy of Sciences, Moscow, U. S. S. R. (See Chapter II, Section A2, Subheading "Academy of Sciences").

b. Special Purpose Computers

            (1) University of Washington Mechanical Translation
                 Pilot Model

                The Electrical Engineering Department constructed a
small, pilot, model translator to demonstrate the technological feasibility
of mechanical translation of languages. It has a very small memory, capable
of storing only 60 German monoglots and their English equivalents. However,
because of the smallness and technological simplicity of the memory device
no German word or English counterpart may contain more than seven letters.
(See Chapter II, Subheading "University of Washington"). In spite of these
severe limitations, and much re-writing of the original text, a crude
mechanical translation was obtained. For example, a pre-edited sample text
used for experimental purposes included the statement, "Aus dem Experiment
kommt oft in kurzer Zeit die Antwort über das, was wahr und was nicht wahr
ist". A fair, close-to-literal translation by a human translator might be,
"Out of the experiment there often comes, in a short time, the answer
regarding that which is true and that which is not true". A free-flowing
translation might be, "By means of an experiment, one may quickly determine
which is true and which is false". The Mechanical Translation Pilot Model
yielded the translation "OUTYOF THETHAT EXPERIMENT COMEZS OFT IN SHORT TIME
THETHAT ANSWER REXOVER THETHAT, WHAT TRUE & WHAT NOT TRUE ISXHAS".

            (2) United States Air Force Mechanical Translator Mark I

                The International Telemeter Corporation constructed
translation equipment under a U. S. Air Force contract. (See Chapter II,
Section A, Subheadings "international Telemeter Corporation" and "U. S.
Air Force"). The present Mark I was built to translate Russian into English.
The bilingual dictionary contains about 40,000-words in their inflected
form. Physically, the dictionary is a 10-inch glass disk carrying 30
million bits in 700 concentric tracks in 0.36 inch wide annulus. The
disk's rotation speed is 1200 rpm. The output current from the photo-
multiplier follows the succession of black and white squares. This signal

is fed into computer-type circuits, and compared with words in the input register. Upon obtaining a match, the subsequent information is delivered to output register. Here information is temporarily stored until delivery to high-speed printer.

        2. Components

        None of the existing general purpose computers are suitable for obtaining even a fair translation of an arbitrarily selected piece of source (foreign) language text. There are many problem areas in the field of translation that require attention before a satisfactory fully automatic translation can be obtained. Problem areas requiring attention include high speed direct reading of printed pages of source language text, high capacity, low access time storage systems, ability to perform logical and arithmetic operations rapidly, and a high speed output system in magnetic tape or printed page form. It is difficult to make quantitative statements of requirements for equipment to perform mechanical translation, since one always arrives at the conclusion that what is desirable is the best translation obtainable at the least possible expenditure of time and money. An estimate of minimum requirements for a suitable, economical translation machine might be the following:

        (a) Character sensing of printed pages of source language text at the rate of the order of thousands of characters per second.

        (b) Storage capacity of the order of not less than tens of millions of bits, with random access time and arithmetical and operation times of the order of not more than hundreds of microseconds.

        (c) Output speed not less than the input character sensing rate.

        Recent advances in computer technology indicate that components with the above capabilities are becoming a reality. The U. S. Air Force Mechanical Translator Mark I approaches some of these capabilities. Additional contracts are being let to further the development of the art in order that the above requirements can be met during the next few years.

In machine translation by digital computers we must execute the following sequence of events:

(a) Human entry of bulk material into the machine

(b) Sensing source language

(c) Converting to a machine code for temporary storage and processing

(d) Translating to target language code by a sequence of logical operations and look-ups

(e) Transfering results to an output medium

(f) Preparing for human consumption

Involved in each of these steps is an information handling rate. Since there can be but little storage at any point in the system, the rate of flow of the system can only be as fast as the rate of flow of the slowest unit, with all other units "waiting" for the slowest. When technological progress permits increasing the information handling rate of the slowest unit, another "slowest" unit becomes the subject for improvement.

It this never-ending spiral toward increased over-all speed, where do we stand today with regard to mechanical translation of foreign languages by general purpose digital computers? Look-up time is the speed limiting operation when a good translation is desirable. Advances in storage techniques may soon remove storage as the limiting element. Since high speed printing and magnetic tape output is available, the conversion of results to an output medium is not a present limiting element. If we admit the possibility of preparing the input in magnetic tape form, input should not become limiting. This would have to be done at the present time by several off-line converters i.e. persons who would convert the printed scientific source language text to magnetic tape. (In order to avoid the problem of having many "converters", a character sensing device could either read directly into the computer or could convert to magnetic tape off-line). This would make the logical operations, required

41

for processing each semantic unit, the limiting factor Most attention is
being devoted to rapid character sensing of the printed page and high speed,
high capacity, random access storage units. The following information merely
indicates the present state of technology of commercially available equipment.

    a. Input Devices

        A large number and variety of computer input devices and
information recording media are available on the commercial market. Many
of these media would be suitable for mechanical translation applications,
were it not for the hard fact that source language text is not recorded on
any of these. It is rather doubtful for instance, that we could entice
Soviet technical societies and organizations into supplying us with their
scientific results on cards or tapes. We must contend, therefore, with
typed or printed pages on which are recorded source language signs and
symbols. Limited progress has been made in automatic reading of printed
matter. Naturally, systems of checking the sensed information must be
developed. In this area checking can be based on the inherent redundancy
of the written source language text, the "impossibility" of certain letter
combinations, and a restricted list of permissible characters.

        Some progress has been made in character sensing using
special prepared ink. This however, is also not a suitable system for
mechanical translation, since foreign publications will probably not be
printed with these special inks. Therefore, a direct-reading character-
sensing device must be developed in order to avoid the use of human readers
and manual input keyboards. Though the human operator need not have any
knowledge of the source language, the eye-brain-hand or the reflexive eye-
hand operation would remain a limiting factor in the mechanical trans-
lation system, since human reflex and muscular action time is slow compared
to machine operation time. (See BRL Report 1010 "A Second Survey of
Domestic Electronic Digital Computing Systems" for description of charac-
teristics of these equipments).

Organizations which have conducted investigations in the field of character sensing include most of the computer manufacturers. Although many organizations have experimented with special inks and dyes, relatively few have investigated character sensing of inactive ink or ordinary paper, such as this page. Among these organizations are Control Instruments Company, a subsidiary of the Burroughs Corporation, the International Business Machines Corporation, Sperry-Rand Corporation, the National Cash Register Company and the National Bureau of Standards, and the Diamond Ordnance Fuze Laboratory. Technical Report No. TR-128, 26 November 1954, Diamond Ordnance Fuze Laboratories, Ordnance Corps, U. S. Army, entitled "DOFL First Reader" describes equipment which accomplished character sensing, Several types of patterns such as maps, diagrams and fingerprints, could be read by the unit. Two commercial organizations have built and delivered operational equipment on a large scale designed to sense printed characters on commercial documents such as gasoline credit cards, travellers checks, telephone bills, dividend disbursement notices, and cash register receipts. These two are:

(l) Intelligent Machines Research Corporation
7019 Edsall Road
Alexandria, Virginia

Six alphanumeric character sensing machines have been delivered and are in operation. These machines are with Standard Oil of California, Socony-Mobiloil, First National City Bank of New York, National Biscuit Company and the Readers Digest Association. Twelve machines will be in operation by the end of 1958. Contracts are in effect with the U. S. Air Force to develop equipment to automatically read 8-1/2 x 11 alphanumeric typed sheets and punch on a high speed paper tape perforator. IMR has developed numerous equipments for automatic sorting of typed-address mail. The city and state typed portion of the envelope is sensed for sorting purposes. No attempt has been made to sense script. The IMR Corporation uses a mechanical-optical system of sensing, requiring a source of illumination, a system of lenses, a system of stationary and

43

rotating slits for generating a scanning beam, photoelectric cells, paper feeding mechanisms, and necessary logical and recording devices. After reducing the character to an established code, the sensed information is temporarily stored and then recorded or transferred to another medium for tabulation or other form of additional processing.

(2) Solartron Electric Group

Solartron, a British firm, is now delivering character recognition devices for reading printed pages and transferring the data to digital computers. The equipment is marketed under the trade name of E. R. A. - Electronic Reading Automaton. E. R. A. scans each letter of print in a series of vertical strokes by a beam of light from a cathode ray tube. The varying reflected light that comes from the paper is reproduced as a varying electric signal by a photocell. Pulse shaping circuitry reduces the "information" to digital data, which is then stored. Logical circuits encode the stored patterns and transfer the data to output lines. The machine can "read" through some blurr and smudge, such as that which might be produced by carbon paper or by overstamping. In short, E. R. A. automatically reads printed data on existing documents and gives a direct translation from printed "human language" to internal "machine language". The first production model has been purchased by the Boots Pure Drug Company, Ltd. of Nottingham, England. It will read the printed sales records produced on rolls by the cash registers in Boots' drug stores. The reading rate is 200-300 character per second. This is approximately equivalent to read-sensing a 8-1/2 x 11 sheet of double spaced typed characters in two seconds. The major difference between the sensing technique used by Solartron and the technique used by the Intelligent Machines Research Corporation is in the method in which the sweeping light beam is generated. Solartron uses an electronic system whereas I. M. R. uses a mechanical one. I. M. R. is ahead in both delivery and reliability, however, Solartron's system reads more rapidly. It must be remembered that the mechanical handling of the document requires more time than the actual sensing, thus

increases in sensing rates does not contribute too greatly to the overall reading rate. Reliability is of great significance, and checking techniques must be built in to the system at the present time. This is done with check characters to insure that the equipment is sensing and interpreting properly.

b. Output Devices

There is a large selection of output devices and techniques for handling the machine output data i.e. the target language text. Translations can be placed on tape for subsequent conversion to printed page. It can be used to actuate high speed printers on-line. Here, a question of economy of operation and system feasibility governs the choice of procedures to adopt. Magnetic tape output, off-line conversion, and high speed printing equipment availability and characteristics is a large and diversified subject and is beyond the scope of this report. (BRL Report 1010 "A Second Survey of Domestic Electronic Digital Computing Systems" contains data on much of this standard equipment).

High speed printers vary considerably in principle of operation, legibility of copy and format control. For instance, for the formation of the printed character, one finds the print wheel, bar, and cylinder solid type, matrix type (usually 5 x 7); cathode ray tube print, and electrochemical, electrostatic and electromagnetic formation of characters. Among the manufacturers of such equipment are:

ACF Electronics
Division of ACF Industries, Inc.
800 Worth Pitt Street
Alexandria, Virginia

Analex Corporation
150 Causeway Street
Boston 14, Massachusetts

Burroughs Corporation
Special Machines Department
219 - 4th Avenue
New York 3, New York

Clary Corporation
San Gabriel, California

Consolidated Vultee Aircraft Corp.
San Diego 12, California

Control Instrument Company
Subsidiary of Burroughs Corporation
67 - 35th Street
Brooklyn 32, New York

Datamatic Corporation
151 Needham Street
Newton Highlands 6l, Massachusetts

Electrodata Division
Burroughs Corporation
460 Sierra Madre Villa
Pasadena, California

Ferranti Electric, Incorporated
30 Rockefeller Plaza
New York 20, New York

Ford Instrument Company
Division of Sperry Rand Corporation
31-10 Thomson Avenue
Long Island City 1, New York

General Dynamics Corporation
Stromberg-Carlson Division
San Diego, California

International Business Machines Corporation
590 Madison Avenue
New York 22, New York

National Cash Register Company
Electronics Division
Hawthorne, California

Potter Instrument Company
115 Cutter Mill Road
Great Neck, New York

Powers - Samas House
Holborn Bars
London, E. C. 1

Remington Rand Univac
Division of Sperry - Rand Corporation
Univac Park
St. Paul 16, Minnesota

Shepard Laboratories, Inc.
480 Morris Avenue
Summit, New Jersey

Underwood Corporation
Electronic Computer Division
55-10 56th Avenue
Long Island City

Printing speeds and legibility of the printed output text varies considerably from printer to printer. In general, obtainable speeds are of the order of 10,000 to 100,000 words per minute of running English text. Fair legibility is obtainable at these speeds. If post-editing of the target language is required another off-line printing step will be necessary. This would mean that the text will be "filtered" by a human operator, who would have to "clean" it up. If the high speed printer output is in usable form, consideration must be given to methods of repro-duction for the distribution of a large number of copies.
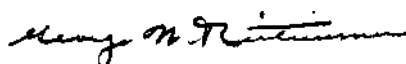
c. Storage Devices

When dealing with storage devices, two major areas of consideration are involved; namely, capacity and access time. The capacity required for the storage of vocabulary and instructions in order to produce an adequate translation in a given scientific field is of the order of 10 to 100 megabits. This amount of storage is available at the present time in the medium of magnetic tape. However, average access time for individual words would require approximately 3 minutes. (Based on 2400 ft. reel running at 75 in/sec). This situation might be improved by placing the most frequent-ly used terms in close proximity to the point of initial search. Other schemes for reducing access time might be accomplished by redundant storage. The utilization of magnetic tape systems and the characteristics and availa-bility of existing magnetic tape systems is too broad a field to be included in this report. At any rate, the limitation in the use of magnetic tape lies not in storage capacity (Approximately 500,000 words/2400 ft. reel and as many as 10 or more handlers) but in the excessive look-up time that would be required. (See BRL Report 1010 for use of magnetic tapes with general purpose electronic computing systems).
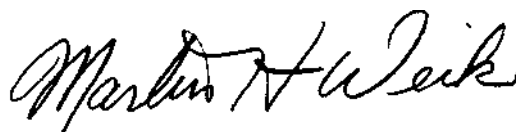
Another alternative lies in the use of magnetic cores, or their companion elements, thin films and ferrite plates. Core units with random access time of the order of a few microseconds are available. Here, however, limitations are imposed by the storage capacities. Capacity limits on existing large scale digital machines are approximately 50,000 words, which is only about one tenth of the requirement for achieving a good translation.

In between the high capacity, slow access tape and the low capacity, fast access core lies the drums, discs, magnetic "cards" and loop-tape "bins". These systems are the results of attempting to obtain both high capacity and fast access. However, both storage capacity and speed were sacrificed. Perhaps these systems are best used as the bulk storage for the least frequently used terms of an automatic dictionary. Examples of some of the large capacity, fast access storage units are given in Table I. Only units capable of storing approximately a million or more words were considered for inclusion in this table.

As an example of the organization of the data in one of the storage systems, the RAM is considered. In order to achieve this combination of large capacity and random access, strips of tape approximately two feet long are strung vertically on stainless steel frames. Each frame is called a page and has five paragraphs. Each paragraph has 20 strips of tape (10 strips mounted back to back). Each tape strip accommodates 20 words (data locations, with 160 alphanumeric characters per data location). Each paragraph contains 400 data locations. Five paragraphs form a RAM page and 200 pages form the RAM book or RAM unit.

GEORGE W. REITWIESNER

MARTIN H. WEIK

49

TABLE I

Computing System	Type of Magnetic* Storage System	Storage Capacity	Average Random Access (Seconds)
RAMAC (IBM)	Disc	600,000 10 Dec Dig Words	0.8 - 60 words
DATAFILE (Electrodata)	Tape Loops	2,000,000 10 Dec Dig Words	24
HD-Magnetic File(LFE)	Drum	15,000,000 bits	0.18
UNIVAC (Sperry-Rand)	Dual Disc	24,000,000 7 Bin Dig Alphanumeric Characters	0.1 to 0.25
LARC (Sperry-Rand)	Drum	1,000,000 12 Dec Dig Words	0.1
LARC (Sperry-Rand)	Disc	2,000,000 12 Dec Dig Words	0.1
RAM (Potter)	Tape Loops	500,000,000 Binary Digits 64,000,000 8 Bin Dig Alphanumeric Characters 400,000 Data Locations of 160 Characters Each	7

*See Chapter II, Subheading Bl, par b(1) for a description of the
 U. S. Air Force Photoscopic Disc Storage Unit.

STORAGE CAPACITY AND AVERAGE RANDOM ACCESS TIME FOR

SOME TYPICAL COMMERCIAL STORAGE SYSTEMS

III BIBLIOGRAPHY IN THE FIELD OP MECHANICAL TRANSLATION OF LANGUAGES

1951

"Studies in Mechanical Translation - No. 1", Erwin Reifler, Department of Far Eastern and Slavic Languages and Literature, University of Washington, Apr. 1951.

"The Present State of Research on Mechanical Translation", Yehoshua Bar-Hillel, American Documentation 2:229-237, 1951.

1952

"Studies in Mechanical Translation", Erwin Reifler, Report on Research Results in Mechanical Translation in Connection with the Rockefeller Grant for the Summer Quarter 1952, University of Washington, Summer 1952.

1953

"Progress Report I", Cambridge Language Research Unit, 20 Millinton Road, Cambridge England, Jan. 1953.

"A Study for the Design of an Automatic Dictionary, Anthony G. Oettinger, Progress Report No. 26, Computation Laboratory, Harvard University, 10 Feb. 1953.

"Mechanical Translation, Andrew D. Booth, Computers and Automation, Vol 2, No. 4, May 1953.

"Machine Translation", Yehoshua Bar-Hillel, Computers and Automation, Jul. 1953.

"The Mechanical Translation of Russian: A Preliminary Report", Kenneth E. Harper, Modern Language Forum, Vol. 38, No. 3-4, pages 12 - 29, Sep. - Dec. 1953.

"An Idioglossary for Mechanical Translation", Victor A. Oswald Jr. and Richard H. Lawson, Modern Language Forum, Vol. 38, No. 3-4 pages 1 - 11, Sep. - Dec. 1953 .

"A Quasi - Arithmetical Notation for Syntactic Description", Yehoshua Bar-Hillel, Language, Vol. 29, pages 47 - 58, 1953.

"An Idioglossary for Mechanical Translation", Victor A. Oswald Jr. and Richard H. Lawson, University of California, Los Angeles, 16 pages, 1953.

"Some Linguistic Problems Connected with Machine Translation", Yehoshua Bar-Hillel, Philosophy of Science, 1953.

"The Mechanical Translation of Russian: A Preliminary Report", Kenneth E. Harper, University of California, Los Angeles, 26 pages, 1953.

1954

"A Short Report on the Present Stage in the Development of Mechanical Translation", Erwin Reifler, Far Eastern Department, University of Washington, Jan. 1954.

"MT, its Psychological Aspects and its Significance for the Human Society", Erwin Reifler, Presented at the Second Conference on General Semantics in St. Louis, 12 Jun. 1954.

"The IBM-Georgetown University of MT Demonstration", Peter Sheridan, Presented at the Meeting of the ACM at Ann Arbor, Michigan, Jun. 1954.

"A Study for the Design of an Automatic Dictionary", Anthony G. Oettinger, Doctoral Theses, Harvard University, Apr. 1954.

"The First Conference on Mechanical Translation", Erwin Reifler, Mechanical Translation, Vol I", No. 2, pages 23 - 32, Aug. 1954.

"The Machine and the Man", Victor H. Yngve, Mechanical Translation, Vol. 1, No. 2, pages 20 - 22, Aug. 1954.

"Computing Machines for Language Translation", T. M. Stout, Mechanical Translation, Vol. 1, No. 3, pages 41 - 46, Dec. 1954.

"The Conference on Mechanical Translation", A. C. Reynolds Jr., Mechanical Translation, Vol. 1, No. 3, pages 47 - 55, Dec. 1954.

"The Distribution of Word Length in Technical Russian", Anthony G. Oettinger, Mechanical Translation, Vol. 1, No. 3, pages 38 - 40, Dec 1954.

"Research in Language Translation on the IBM Type 701", Peter Sheridan, IBM Applied Science Division Technical Newsletter, No. 9, pages 5 - 24, Jan. 1955.

"Zhurnal Experimental'noi I Teoretichesk'oi Fiziki", Vol. 26, No. 2, pages 189 - 207, Feb. 1955.

"Machine Translation of Languages", William N. Locke and Donald A. Booth, The Technology Press of MIT and John Wiley and Sons, Inc., New York, Chapman Hall, Ltd., London, May 1955.

"Mechanical Determination of the Constituents of German Substantive Compounds", Erwin Reifler, Mechanical Translation, Vol. 2, No. 1, pages 3 - 14, Jul. 1955.

"Translation of Russian Technical Literature by Machine", James W. Perry, Mechanical Translation, Vol. 2, No. 1, pages 15 - 24, Jul. 1955.

"Machines for the Translation of Languages", Victor H. Yngve, Journal of Communication, Vol. 5, No. 2, pages 35 - 40, Summer 1955.

"Translating Russian by Machine", Kenneth E. Harper, Journal of Communication, Vol. 5, No. 2, pages 4l - 46, Summer 1955.

"The Translation of Languages by Machine", Victor H. Yngve, Research Laboratory for Electronics - MIT, 1955.

"A Statistical Structural Analysis of a Model Language", Frederick P. Brooks, Papers Presented at the Seminar in Mathematical Linguistics, (Vol. 1) Harvard University, 1955.

"An Automatic Glossary and Problems of Case Translation", Henry Landau, Papers Presented at the Seminar in Mathematic Linguistics, (Vol, l) Harvard University, 1955.

"Problems in the Mechanical Translation of Chinese", John Young, Papers Presented to the Seminar in Mathematical Linguistics, (Vol. l), Harvard University, 1955.

"The Design of an Automatic Russian-English Technical Dictionary", Anthony Oettinger, Chapter 3 in Machine Translation of Languages, W. N. Locke and A. D. Booth (editors) John Wiley and Sons, New York, 1955.

1956

"The UW Automatic Language Translator", Geoff Douthwaite, Washington Engineer, Vol. 8, No. 4, Feb. 1956.

"Can Translation be Mechanized?" Yehoshua Bar-Hillel, Mada, Vol. 1, No. 2, Apr. 1956. (In Hebrew)

"A Machine Translates from one Language to Another", Lew R. Mickleson, The Trend, Vol. 8, No. 3, Jul. 1956.

"Engineering Process in Machine Translation", Robert E. Wall, Jr., The Trend, Vol. 8, No. 3, Jul. 1956.

"Gap Analysis and Syntax", IRE Transactions on Information Theory, Vol. 1T-2 No. 3, pages 106 - 112, Sep. 1956.

"Three Models for the Description of Language", Noam Chomsky, IRE Transaction on Information Theory, Vol. IT-2, No. 3, pages 113 - 124, Sep. 1956.

"Mecanismes et Traduction", Edmund Cary, Babel, Vol. 2, No. 3, pages 102 - 107, Oct. 1956.

"Present Objectives of MT Research in the United Kingdom", Andrew D. Booth, Babel, Vol. 2, No. 3, pages 108 - 110, Oct. 1956.

"Previous Experiments in Mechanical Translation", L. Brandwood, Babel, Vol. 2 No. 3, pages 125 - 127, Oct. 1956.

"Teaching the Machine Grammar", M. V. Heberden, Babel, Vol. 2, No. 3, pages 119 - 124, Oct 1956.

"Terminology in the Light of Research on Mechanical Translation", V. H. Yngve, Babel, Vol. 2, No. 3, pages 128 - 132, Oct. 1956.

"The Outlook for Mechanical Translation", Victor H. Yngve, Babel, Vol. 2, No. 3, pages 99 - 101, Oct. 1956.

"The Outlook for Mechanical Translation", Victor H. Yngve, Le Linguiste, No. 4, pages 5-7, 1956 (Reprinted in Babel, Vol 2, No. 3, pages 99 - 101, Oct. 1956.

"The Translation of Foreign Language by Machine", L. Brandwood, Babel, Vol. 2, No. 3, pages 111- 118, Oct. 1956.

"Contextual Analysis in Word-for-word MT", Kenneth E. Harper, Mechanical Translation, Vol. 3, No. 2, pages 40 - 4l, Nov. 1956.

"Mechanical Translation Work at the University of Michigan", A Koutsoudas and R. Machol, Mechanical Translation, Vol. 3, No. 2, Nov. 1956.

"Machine Translation, Development at the University of Washington", Erwin Reifler, Mechanical Translation, Vol. 3, No. 2, Nov. 1956.

"Mechanical Translation and the Problem of Multiple Meaning", A. Koutsoudas and R. Korfhage, Mechanical Translation, Vol. 3, No. 2, pages 46 - 51, Nov. 1956.

"Mechanical Translation of French", L. Brandwood, Mechanical Translation, Vol. 3, No. 2, pages 52 - 6l, Nov. 1956.

"Mechanical Translation Research at MIT", Victor H. Yngve, Machine Translation, Vol. 3, No. 2, pages 44 - 45, Nov. 1956.

"On the Problem of Mechanical Translation", D. Panov (Translation from Russian by M. Friedman and M. Halle) Mechanical Translation, Vol. 3, No. 2, pages 42 - 43, Nov. 1956.

"Organization and Method in Mechanical Translation Work", L. E. Dostert, Mechanical Translation, Vol. 3, No. 2, Nov. 1956.

"Some of the Engineering Aspects of the Machine Translation of Language", Robert E. Wall, Jr., Communication and Electronics, AIEE, No. 27, pages 580 - 585, No. 1956.

"Stochastic Methods of Mechanical Translation", Gilbert W. King, Mechanical Translation, Vol. 3, No. 2, pages 38 - 39, Nov. 1956.

"The Computer as a Translator", A Koutsoudas and R. Korfhage, <u>Michigan Alumnus</u>, Vol. 63, No. 10, pages 34 - 37, Dec. 1956.

"A General Program for Mechanical Translation between any two Languages via an Algebraic Interlingua", R. H. Richens, <u>Cambridge Language Research Unit</u>, Cambridge, England, 1956.

"The Linguistic Basis of the Thesaurus - Type Mechanical Dictionary and its Application to English Preposition Classification", M. A. K. Holliday, <u>Cambridge Language Research Unit</u>, Cambridge, England, 1956.

"Adamo II", by Enrico Maretti, <u>Civilta delle macchine</u>, No. 3, pages 25 - 32, 1956.

"An Algebraic Thesaurus", A. F. Parker - Rhodes <u>Cambridge Language Research Unit</u>, Cambridge, England, 1956.

"Automatic Translation", D. Y. Panov, <u>Academy of Sciences of the USSR</u>, <u>Moscow</u>, 45 pages, in Russian, 1956.

"Basic Problems of Machine Translation", P. S. Kuznetsov, A. A. Lyapunov, and A. A. Reformatskiy, VYa, No. 5, pages 109 - HO, 1956.

"Certain Problems of Automatic Translation", (Air Technical Intelligence Translation), I. K. Bel'skaya, L. N. Korolev, I. S. Mukhin, D.Yu Panov, and S. N. Razumooskiy, <u>Vestnik Academii Nauk SSSR</u>, Vol. 26, No. 12, pages 24 - 33, ATIC - 306672 - F - TS - 9118/V, 1956.

"Concerning the Problem of Machine Translation of Languages", D. Panov, <u>Academy of Sciences</u>, <u>USSR</u>, 35 pages, 1956.

"Form Classes: Structural Linguistics and Mechanical Translation", Lew R. Micklesen, <u>For Roman Jacobson,</u> Mouton and Company, The Hague, pages 344 - 352, 1956.

"Linguistics Analysis and Translation", J. R. Firth, <u>For Roman Jacobson</u>, Mouton and Co., The Hague, pages 133 - 139, 1956.

"An Electronic Computer Program for Translating Chinese into English" A. F. Parker - Rhodes, <u>Mechanical Translation</u>, Vol. 3, No. 1, pages 14 - 19.

"Graphic Linguistics and its Terminology", R. A. Crossland, <u>Mechanical Translation,</u> Vol. 3, No. 1, pages 8 - 11, 1956.

"Preprogramming for Mechanical Translation", R. H. Richens, <u>Mechanical Translation</u>, Vol. 3, No. 1, pages 20 - 25, 1956.

"Progress Report II", <u>Mechanical Translation</u>, Vol. 3, No. 1, 1956.

"Some Linguistics Problems in Machine Translation", Paul Garvin, <u>For Roman Jacobson</u>, Mouton Company, The Hague, pages 18o - 186, 1956.

"The Potentialities *of* a Mechanical Thesaurus", Margaret Masterman, <u>Cambridge Language Research Unit</u>, Cambridge, England, 1956.

"The Technical Feasibility of Translating Languages by Machine, Victor H. Yngve, <u>Electrical Engineering</u>, Vol. 75, No. 11, pages 994 - 999, 1956.

"The Translation of Languages by Machine", Victor H. Yngve, <u>Information Theory</u>, (Third London Symposium) Butterworth's Scientific Publications, London, pages 195 - 205, 1956.

"A Method for Using Punctuation Patterns in the Machine Translation of Languages", Gerard Salton, <u>Papers Presented to the Seminar in Mathematical Linguistics</u>, (Vol II) Harvard University, 1956.

"A Preliminary Study of Contextual Constraint on Multiple Correspondents in a Russian - English Automatic Dictionary", Stephen Warshall, <u>Papers Presented to the Seminar in Mathematical Linguistics</u>, (Vol. II) Harvard University, 1956.

"A Study of Some Problems of Structural Ambiguity in Machine, Translation of Languages", Richard Karp, <u>Papers Presented at the Seminar in Mathematical Linguistics</u>, (Vol.II) Harvard University, 1956.

"Mechanical Resolution of Syntax Patterns: Hebrew", Harvey Sobelman, <u>Papers Presented to the Seminar in Mathematical Linguistics</u> (Vol. II) Harvard University, 1956.

"Multiple Correspondence in Automatic Translation", Roderick Gould, <u>Papers Presented at the Seminar in Mathematical Linguistics</u> (Vol. II) Harvard University, 1956.

"The Implications of Several New Methods of Sentence Analysis for Machine Translation", Hubert Hocutt, <u>Papers</u> <u>Presented</u> <u>at</u> <u>the</u> <u>Seminar</u> <u>in</u> <u>Mathematical</u> <u>Linguistics</u>, (Vol. II) Harvard University, 1956.

<center>1957</center>

"An Automatic Method of Inflection for Russian", Karol Magassy, <u>Design</u> <u>and</u> <u>Operation</u> <u>of</u> <u>Digital</u> <u>Calculating</u> <u>Machinery</u>, <u>Progress</u> <u>Report</u> No. 46, Harvard Computation Laboratory, Mar. 1957.

"Noun Phrase Analysis using the IBM 650 Computer", Aristotelis D. Stathacopoulos, <u>The</u> <u>Trend</u> <u>in</u> <u>Engineering</u>, University of Washington, Apr. 1956.

"Frequency of Occurrence of Words: A Study of Zipf's Law with Application to MT", <u>University</u> <u>of</u> <u>Michigan</u>, External Report 2144-147-T, June 1957.

"Mechanical Translations", Andrew D. Booth, <u>ASLIB</u> <u>Proceedings</u>, Vol. *9,* No. 6, pages 177 - 181, June 1957.

"Conference on Problems of Development and Construction of Information Machines", I. A. Mel'chuk, <u>Voprosy</u> <u>Yazykoznaniya</u>, No. 5, pages 161 - 162, Sep. - Oct. 1957.

"Machine Translation of Languages", I. K. Belskaja, (Academy of Sciences, USSR),. <u>Research</u> - <u>Butterworth's</u> <u>Scientific</u> <u>Publications</u>, Vol. 10, No. 10, pages 383 - 389, Oct. 1957.

"Studies in Machine Translation - 2: Research Methodology", H. P. Edmundson, D. G. Hays, P-1251, <u>The</u> <u>Rand</u> <u>Corporation</u>, Santa Monica, California, Dec. 1957.

"A Stochastic Approach to a Linguistic Problem", A. Koutsoudas and A. Humecky, <u>University</u> <u>of</u> <u>Michigan,</u> 1957.

"Experiments in Machine Translation", T. W. Moloshnaya, <u>Voprosy</u> <u>Yazykoznaniya</u>, No. 4, pages **928-9?97,** 1957.

"Order of Subject and Predicate in Scientific Russian", I. Lehiste, <u>University</u> <u>of</u> <u>Michigan</u>, 1957.

"Papers Presented at the Seminar in Mathematical Linguistics", Widener Library, Vols. 1-2, 1955 - 1957.

"Case Endings of Nouns in Russian", Mr. Pacak, Georgetown University, Machine Translation Project, Seminar Work Paper, No. MT-15, 1957.

"Word Analysis Versus Multiple-Entry Glossary Storage" Mr. Brown, Georgetown University, Machine Translation Project, Seminar Work Paper No. MT-18, 1957.

"Verbal Items and their Glossarization", Mr. Pacak, Georgetown University, Machine Translation Project, Seminar Work Paper No. MT-19, 1957.

"Review of Noun, Adjective and Verb Suffixes", Mr. Pacak, Georgetown University, Machine Translation Project, Seminar Work Paper No. MT-20, 1957.

"A Suggested Method for Achieving MT with a Short Trial on Chemical French" Mr. Brown, Georgetown University, Machine Translation Project, Seminar Work Paper No. MT-21, 1957.

"The Range of Machine Search for Translation of Russian Pronouns", Mr. Zarechnak and Miss Pyne, Georgetown University, Machine Translation **Project,** Seminar Work Paper No. MT-22, 1957.

"Double Translation", Mr. Pantzer, Georgetown University, Machine Translation Project, Seminar Work Paper No. MT-24, 1957.

"Texts and the Translation Unit", Mr. Pantzer, Georgetown University, Machine Translation Project, Seminar Work Paper No. MT-25, 1957.

"Preliminary Report on the Article", Mr. Glazer, Georgetown University, Machine Translation Project, Seminar Work Paper No. MT-26-, 1957.

"Russian Phraseological Expressions", Mr. Sushko, Georgetown University, Machine Translation Project, Seminar Work Paper No. MT-27,, 1957.

"The Methodology of the Experimental Group", Miss Pyne and Experimental Group, Georgetown University, Machine Translation Project, Seminar Work Paper, No. MT-28, 1957.

"Basic Syntactic Concepts", Mr. Zarechnak, <u>Georgetown</u> <u>University</u>, <u>Machine</u> <u>Translation</u> <u>Project</u>, Seminar Work Paper Wo. MT-29, 1957.

"Three Russian Prepositions", Miss Fargo and Miss Rubin, <u>Georgetown</u> <u>University</u>, <u>Machine</u> <u>Translation</u> <u>Project</u>, Seminar Work Paper No. MT-30, 1957.

"The Russian Preposition 'K' ", Miss Fargo and Miss Rubin, <u>Georgetown</u> <u>University</u>, <u>Machine</u> <u>Translation</u> <u>Project</u>, Seminar Work Paper No. MT-31, 1957.

"Identification of Russian Items by the Machine Procedure", Mr. Pacak, <u>Georgetown</u> <u>University</u>, <u>Machine</u> <u>Translation</u> <u>Project</u>, Seminar Work Paper No. MT-32, 1957.

"Cue Lists for Nominal and Adjectival Suffixes", Miss Mathoit and Experimental Group, <u>Georgetown</u> <u>University</u>, <u>Machine</u> <u>Translation</u> <u>Project</u>, Seminar Work Paper No. MT-33, 1957.

"General Flow Chart for MT", Mr. Belmore, <u>Georgetown</u> <u>University</u>, <u>Machine</u> <u>Translation</u> <u>Project</u>, Seminar Work Paper No. MT-34, 1957.

"Statement of Proposed Method for Mechanical Translation", Miss Lukjanow, <u>Georgetown</u> <u>University</u>, <u>Machine</u> <u>Translation</u> <u>Project</u>, Seminar Work Paper No. MT-35, 1957.

"The Theoretical Basis of the Uni-Directional Matching Method", Mr. Austin <u>Georgetown</u> <u>University</u>, <u>Machine</u> <u>Translation</u> <u>Project</u>, Seminar Work Paper No. MT-36, 1957.

"Pronominal 'Shto' ", Miss Fargo and Miss Rubin, <u>Georgetown</u> <u>University</u>, <u>Machine</u> <u>Translation</u> <u>Project</u>, Seminar Work Paper No. MT-37, 1957.

"English Syntactic Concepts for MT", Miss Pyne, <u>Georgetown</u> <u>University</u>, <u>Machine</u> <u>Translation</u> <u>Project</u>, Seminar Work Paper No. MT-38, 1957.

"Tentative Statement for Choice in the Translation of Noun Suffixes", Miss Fargo and Miss Rubin, <u>Georgetown</u> <u>University</u>, <u>Machine</u> <u>Translation</u> <u>Project</u>, Seminar Work Paper No. MT-40, 1957.

"Preliminary Materials for Coding", Mr. Garvin and Experimental Group, <u>Georgetown</u> <u>University</u>, <u>Machine</u> <u>Translation</u> <u>Project</u>, Seminar Work Paper No. MT-41, 1957.

"Article Requirements of Plural Nouns in Russian Chemistry Texts", Mr. Glazer, Georgetown University, Machine Translation Project, Seminar Work Paper No. MT-42, 1957.

"Formulation (The work which must precede the programming of MT)", Mr. Belmore, Georgetown University, Machine Translation Project, Seminar Work Paper No. MT-47, 1957.

"Syntactic Transfer Procedures", Mr. Zarechnak and Miss Pyne, Georgetown University, Machine Translation Project, Seminar Work Paper No. MT-48, 1957.

"The Transfer of Russian Reflexive verbs into English", Mr. Pacak, Georgetown University, Machine Translation Project, Seminar Work Paper No. MT-50, 1957.

"Routines Involving Agreement", Mr. Garvin and Experimental Group, Georgetown University, Machine Translation Project, Seminar Work Paper No. MT-52, 1957.

"A Linear System of English Syntax", Mr. Austin, Georgetown University, Machine Translation Project, Seminar Work Paper No. MT-53, 1957.

"Impersonal and Infinitive Structures in Russian and their Transfer into English", Mr. Pacak, Georgetown University, Machine Translation Project, Seminar Work Paper No. MT-54, 1957.

"A Preliminary Mechanization of the Method for Machine Translation Suggested in Paper MT-21", Mr. Brown, Georgetown University, Machine Translation Project, Seminar Work Paper No. MT-55; 1957.

"Preposition 'C' ", Miss Nancy Fargo and Miss Joan Rubin, Georgetown University, Machine Translation Project, Seminar Work Paper No. MT-57, 1957.

"An Arrangement Routine Involving Certain Participles (Preliminary Statement)", Mr. Garvin Group II, Georgetown University, Machine Translation Project, Seminar Work Paper No. MT-60, 1957.

"Routines and Loops in Syntactic Searching", Mr. Garvin Group II, Georgetown University, Machine Translation Project, Seminar Work Paper No. MT-61, 1957.

"Routines and Loops in Syntactic Searching and Rearrangement", Mr. Garvin Group II, <u>Georgetown University</u>, <u>Machine Translation Project</u>, Seminar Work Paper No. MT-62, 1957.

"An Automatic Method of Inflection for Russian", Karol Magassy, <u>Design and Operation of Digital Calculating Machinery,</u> Progress Report AF-46, Sec. V, Harvard Computation Laboratory, 1957.

"An Automatic Method of Inflection for Russian (II)", Karol Magassy, <u>Design and Operation of Digital Calculating Machinery</u>, Progress Report AF-49, Sec. III, Harvard Computation Laboratory, 1957.

"A Study of Contexts of the Russian Preposition V", William Foust, <u>Papers Presented at the Seminar in Mathematical Linguistics</u>, (Vol. III) Harvard University, 1957.

"Automatic Machines", Anthony Oettinger, <u>Design and Operation of Digital Calculating Machinery,</u> Progress Report AF-49, Sec. II, Harvard Computation Laboratory, 1957.

"Article on 'Translation', 'Mechanical Translation'", Anthony Oettinger, Section to appear in the <u>Encyclopedia Britannica,</u> 1957.

"Automatic (Transference, Translation, Remittance, Shunting)", Anthony Oettinger, Reuben A. Brower, Editor, <u>Harvard University Press,</u> 1957.

"Automatic Translation", Anthony Oettinger, <u>Harvard Monograph in Applied Science</u>, Harvard University Press, 1957.

"Compilation of an Automatic Dictionary", Vincent Giuliano, <u>Design and Operation of Digital Calculating Machinery</u>, Progress Report AF-49, Harvard Computation Laboratory, 1957.

"Linguistic Data Needed in Machine Translation", William Foust, <u>Honors Thesis in Linguistics and Applied Mathematics</u>, Harvard University, 1957.

"Inflected Form Generators", William Foust, <u>Design and Operation</u> of <u>Digital Calculating Machinery</u>, Progress Report AF-49, Harvard University, 1957.

"Linguistics and Mathematics", Anthony Oettinger, <u>Studies Presented to Joshua Whatmough</u>, (E. Pulgram, Editor, Mouton and Co.,'s-Gravenhage, pages 179 - 186, 1957.

"Programming an Automatic Dictionary", Vincent Giuliano, <u>Design</u> <u>and</u> <u>Operation</u> <u>of</u> <u>Digital</u> <u>Calculating</u> <u>Machinery</u>, Progress Report AF-49, Sec. I, Harvard Computation Laboratory, 1957.

"Report on the "Eighth Annual Round Table Meeting on Linguistics and Language Study-Research in Machine Translation", Leon Dostert, <u>Monograph</u> <u>Series</u>, <u>Georgetown</u> <u>University</u>, No. 10, 193 pages, 1957. The following papers were presented:

"Brief Review of the History of Machine Translation Research"
Leon E. Dostert, Georgetown University.

"Meaning in Relation to Machine Translation"
Martin Joos, University of Wisconsin.

"Linguistic Analysis and Mechanical Formulation"
A. C. Reynolds, Stromberg Time Corporation.

"Linguistic Analysis and Translation Analysis"
Paul L. Garvin, Georgetown University.

"Language as Symbolic Logic"
William M. Austin, Georgetown University.

"The Mathematics and Logic of Language"
Duncan Harkin, U. S. Government Consultant

"Character Sensing as in Input to Machine Translation"
Clyde C. Heasly, Jr. Intelligent Machines Research Corporation.

"The Rationale of the Micro-Glossary Techniques"
Victor A. Oswald, U. C. L. A.

"Word Decomposition for Machine Translation"
R. Richens, M. A. K. Halliday, R. A. Crossland, Cambridge Language Research Unit.

"The Problem of Lexical Storage"
Gilbert King, International Telemeter Corporation

"New Large Capacity Optic Data Storage Device"
Jacob Rabinow, Rabinow Engineering Company.

"Linguistic and Information Retrieval"
 Simon M. Newman, U. S. Patent Office.

"Structure of Noun Phrases in German"
 W. P. Lehman, University of Texas.

"Types of Russian Sentences"
 Michael Zarechnak, Georgetown University.

"Some Ideas on Inter-structural Syntax"
 Jane A. Pyne, Georgetown University.

"The Use of the SEAC in Syntactic Analysis"
 Richard B. Thomas, National Bureau of Standards.

"Programming Aspects of Machine Translation"
 Daniel A. Belmore, Consultant Programmer, Washington, D. C.

"Practical Objectives in Machine Translation Research"
 Leon E. Dostert, Georgetown University.

"Selection and Classification of the Vocabulary for the Automatic Dictionary",
Ladislav Matejka, Design and Operation of Digital Calculating Machinery,
Progress Report AF-49, Sec. 4, Harvard Computation Laboratory, 1957.

"Some Psychological Methods for Evaluating the Quality of Translations",
George A. Miller, and J. G. Beebe-Center, Mechanical Translation No. 3,
pages 73 - 80, 1957.

                                1958

"Russian to English Machine Translation", Robert E. Wall and Udo K. Niehaus,
Communication and Electronics, AIEE, No. 34, pages 709 - 714, Jan. 1958.

"Mechanical Resolution of Linguistic Problems", A. D. Booth, J. P. Cleave
and L. Brandwood, Academic Press, 1958.

"Transfer of Participles, Russian to English (Full Forms Only)", J. Pyne,
M. Zarechnak, M. Pacak, Georgetown University, Machine Translation Project,
Seminar Work Paper MT-65, 1958.

"Transfer of Adverbs: Russian - English", J. Pyne, M. Sushko, M. Zarechnak, Georgetown University, Machine Translation Project, Seminar Work Paper MT-66, 1958.

"Transfer of Adverbs: Russian - English", J. Pyne, M. Sushko, M. Zarechnak, Georgetown University, Machine Translation Project, Seminar Work Paper MT-67, 1958.

"Details of a Method for Coding the Mechanical Translation of French into English", A. F. R. Brown, Georgetown University, Machine Translation Project, Seminar Work Paper MT-70, 1958.

"Report of Group II of the GU Project in MT Research", Paul L. Garvin, Dorita Lochak, Madeleine Mathiot, Christine Anne Montgomery, Georgetown University, Machine Translation Project, Seminar Work Paper MT-73, 1958.

"Scheme of Russian Morphology in terms of Mechanical Translation, Dr. Milos Pacak, Georgetown University, Machine Translation Project, 1958.

"Further Details on the Coding of Dictionary Items for the Mechanical Translation of French", A. F. R. Brown, Georgetown University, Machine Translation Project, Seminar Work Paper MT-75, 1958.

"Organization of the Mechanical Dictionary for Machine Translation of French into English", A. F. R. Brown, Georgetown University, Machine Translation Project, Seminar Work Paper MT-76-,1958.

"The Problem of Digit Numbers in Machine Translation", Group III, Georgetown University, Machine Translation Project, Seminar Work Paper Series B, No. 1, 1958.