Technical Note 60-25
December 1960
Revised January 1961

THE DOD PROBLEM IN THE MECHANICAL TRANSLATION FIELD

S. J. Deitchman

Institute for Defense Analyses
Research and Engineering Support Division
Contract SD-50
Task Order T6

CONTENTS

I. INTRODUCTION

Machine translation of languages (MT) has grown from a suggested idea in 1949 to a current effort involving some two and a half million dollars per year spread over thirteen contracts sponsored by the National Science Foundation, Central Intelligence Agency, and the three military services. Many related efforts in applicable theoretical study and equipment development exist. Typically, with some successful results claimed to be in the offing, pressures have been building for transformation of individual research programs into developments leading to routine production of machine translations at a high rate. Two proposals for such production centers - by Georgetown University and the International Business Machines Corporation - have been made. Hearings, in 1960, of a special House subcommittee have resulted in recommendations for a national mechanical translation center,(1)* and a National Academy of Language Sciences which may include such a center.

Despite these external signs of growing maturity in the field, deeper investigation shows that it is in a state of turmoil. There exist in the DOD community many gaps, ambiguities, and contradictions in expressions of requirements and assessment of the capabilities of various programs to meet requirements. Further, these expressions within that community differ from similar expressions in other quarters having an interest in MT.

* Superscript numbers refer to References, pages 41-43.

The purpose of this report, prepared* for the Assistant Director (Special Projects), ODDR&E, is to recommend a course of action for the DOD that will permit resolution of these uncertainties, establishment of a policy with respect to MT, and a program to implement it.

* By the Research and Engineering Support Division, IDA, under Contract No. SD-50, Task Order T6.

II. SUMMARY, ASSESSMENT AND RECOMMENDATIONS

A. Requirements and Program Implementation

Status. It has been found that needs for translation, to which MT may be applicable, vary among and within independently acting agencies of the using community. The CIA believes that it has stated its requirements in this area unequivocally.* It has planned in 1956 and has since been carrying out an MT program at Georgetown University to meet these requirements, and anticipates early success in applying this program to its needs.

The DOD situation is more complex. Among the various DOD agencies concerned, there is general agreement that the needs are distributed among intelligence, scientific/technical, and other more specialized applications. It is also agreed that within these three areas the spectrum of needs varies from screening of roughly translated to preparation and detailed analysis of carefully translated documents. Russian and Chinese are universally accepted as languages of prime importance for MT effort, with other language requirements varying according to user and application.

Few of the various expressed DOD requirements have been documented officially. There is no agreement on the kind, or quality, of translation compatible with or suitable for applications in various parts of

---

* E.g., in References (1) and (2). Since such statements by CIA personnel are reviewed and approved in-house prior to issuance, they are considered official statements of policy.

the spectrum. Various expressions of the quantity or volume of translation required diverge widely, it being apparent only that the demand far exceeds current manual translation capability and will grow. Nor has there been any serious attempt to show how volume requirements are distributed among the various applications and associated requirements for quality. Nevertheless, actions are being taken unilaterally by various DOD agencies to continue, increase, or decrease support of various programs. No logically planned program framework to guide these actions exists within the Department as a whole.

It has become apparent in the course of this study that establishment of such a framework must take place on a coordinated basis for the entire Department of Defense. Many members of the using community have expressed the belief that this job cannot and should not be done unilaterally by one member of the community, whether by individual initiative or assignment. Further, the CIA needs represent a great part of the requirement for MT research and development, and although its financial support of work in the area is relatively small, its program will have a large impact on the general status of the field. It appears therefore that the DOD and CIA must work in cooperation to resolve inter-agency questions that currently exist.

Recommendation. It is recommended that a joint DOD/CIA working group be established with full responsibility to express the community requirements in MT and with authority to act as the DOD/CIA agent to meet those requirements. The group, while it should be small to be

viable, should include one member from the CIA and each interested DOD agency; care should be taken, if it is the DOD/CIA intent to disseminate translated technical and scientific literature to the technical/scientific community, that that community is represented (probably through the National Science Foundation). Each agency representative should have the responsibility for assembling the requirements of various parts of his agency and for representing their interests in implementing a program. The group should exist on a continuing basis and should anticipate extended periods of full-time work as necessary.

It should be noted that two inter-agency committees exist in the MT area: the Subcommittee on MT of the Committee on Documentation, Intelligence Board (CODIB); and the Interagency Committee on MT Research of the NSF. A separate group is recommended here because:

(a) The CODIB group represents only a part of the community interest in MT discussed in this report (see pp. 14 and 15); and

(b) The NSF is not an operating agency.

The primary point of concern is that, for DOD purposes, authority and responsibility could be discharged best if centralized in an appropriate, generally representative, body.

The immediate tasks of the joint group should be the following:

1. A statement should be prepared describing community needs in the three areas of intelligence, dissemination of scientific/technical information, and specialized applications. This would include:
   a. A precise statement of what must be done in the areas of screening, scanning, and detailed study involving both rough and careful translation.
   b. A statement of the quality of translation needed for each application, with some definition. (Since it is apparent that further experimentation with various MT outputs is needed to define better the mechanisms and circumstances of meaning transfer between two languages, plans should be made for carrying out this experimentation.)
   c. A statement of how much of each kind of machine translation is required.*
2. Needs and desires within and between services and agencies should be compared, leading to a single (however complex) DOD/CIA position in each area of requirements.

---

* It may be found early that the community desires all the translation, of each kind, that it can get up to the limits of capability of a given number of machines (see pp. 16 and 17), and that gradual improvement in quality is both desirable and anticipated for all applications. If so, the recognition, statement and support by analysis of these facts will have obvious positive value in guiding DOD efforts.

3. A technical evaluation should be made of the various MT programs, leading to selection of appropriate efforts (if any exist) for immediate application to production-type outputs on an experimental basis in each of the requirements areas, and selection (or initiation) of longer-range research and development efforts for continued (or perhaps increased) support.

4. The requirements position and the technical program plan should be documented.

The work recommended in the following paragraphs will assist the group in carrying out the above tasks.

B. Technical Assessment

Status. As in any technical field, there can be expected in MT an indefinitely continuing period of research which will contribute, from time to time, to development of "production" techniques in translation and to applications of MT as a tool or source of basic knowledge and technique in broader areas related to the general information processing problem. The great majority of MT efforts today are admittedly research efforts. About the few that are claimed to be ready for application to volume output in one way or another, controversy rages.

The core of this controversy lies in the various sponsors' and technical people's assessments of the state of readiness of particular

programs for such application; their assessment of the utility of outputs that are admittedly primitive or need human doctoring to be intelligible; and their judgment of the potential for improvement to meet whatever "standards of goodness" may be established. The various arguments cannot be interpreted by the person technically uninformed in the field, because they involve questions of linguistic theory and computer programming that have to be examined in great detail. Resolution of the arguments will most likely also require the results of the experiments mentioned in 1.b., page 4.

Recommendation. An impartial technical evaluation of the various MT programs should be carried out with the following objectives:

1. For the efforts claimed to be near production capability:
   a. To describe the nature of the output, especially the language distortions that are peculiar to it, and to relate these output characteristics to the underlying machine program and linguistic theory.
   b. To assess the development potential of the respective programs, indicating the probable limitations of refinement, difficulties of programming new languages, and the time expected to reach the most advanced potential development level.
2. For the other research efforts:
   a. To describe them and separate them into two classes - those which are ultimately intended to join the

development/production group, and those which are intended to continue as research programs in automatic language translation or linguistics.

b. To describe and analyze the first group as in 1 above.

c. To describe possible payoffs from the second group - what (in computer programming, intermediate language, resolution of polysemy, etc.) might be learned from them applicable to MT technique, linguistic theory, and general information storage and retrieval problems over the short and long term.

3. The above results for all the programs should be described in such a way that areas of overlap, similarity, difference and omission become apparent, in each portion of the translation process (see page 29).

C. Operational Analysis

The Problem. In all of the investigation described herein, the thesis has been accepted provisionally that the demand for translation is and will be so great that translation by automatic means is the only solution. The validity of this assumption has yet to be examined. It is quite apparent that the creation of a machine translation capability will require large expenditures of effort and manpower, and that the results will inevitably represent various compromises with the

most desirable output. Certainly, if foreign documents could be read by the user or translated manually by those knowledgeable in the language and subject field of the document, many of these compromises would not be necessary. Further, the nature of the compromises is not really clear, in terms of the man's work, the machine's work, and the effects of various interactions on their joint effort.

Whether the volume of output anticipated will necessarily transcend manual translation capability depends, of course, on the number of translators available. The costs (for Russian alone) of training translators and operating a manual facility (with salaries that will attract people into the business) and those of developing and operating an MT facility, including automatic reader, translator, output printer, trained post-editors (if they are needed), and various overhead and support personnel, have not been explored as thoroughly from the system aspect, as they might be.* Nor has it been established that an MT capability will necessarily improve our foreign-document information handling capability, or reduce waiting time for outputs,** in comparison with an expanded manual capability. While it could not be argued that research in MT should not continue for whatever benefits may be obtained,

---

* The CIA, with its considerable experience in the translation and documentation fields, has accumulated comparative cost figures which should prove very useful in such analysis.

** Because, if demand grows with service capacity (as anticipated, cf. p. 17) the two may retain such proportion (if, for budgetary reasons, the necessary number of computers cannot be made available) that the waiting time remains constant or even increases.

the nature of a development and production facility program will certainly be affected by the answers to these questions.

Recommendation. An operational analysis of the manual and machine translation processes should be carried out, with the following objectives:

1. to delineate for each method the steps in the translation process;*
2. to describe the time and effort distribution of these steps;
3. to delineate the roles and required qualifications of various personnel and describe their tasks for each method; and
4. to estimate the cost in dollars, equipment, and personnel, including research, development, operation and training (as applicable), for given volumes of translation output by each method.

D. Choice of Contractor

It is apparent from the results reported herein that a relatively large amount of effort by persons with specialized knowledge in the fields of languages, computer programming and operation, operational analysis, and information storage and retrieval will be necessary to carry out the technical evaluation and operational analysis recommended above. This suggests that the DOD must fund an external effort for this purpose. A single contractor is recommended for this effort because of the close relationship between the two areas of study.

* This step would clearly borrow some results from the technical evaluation recommended on page 6.

The results reported herein indicate that it will be very difficult to find a potential contractor, knowledgeable in the MT field, who has no commitment to one of the many approaches to the problem. Preservation of the necessary objectivity is thus difficult.

It is suggested that this question be resolved by selection of a contractor for the tasks who has the general qualifications outlined above, but who has not been active in the MT field. This implies that a certain part of the cost of the study will be required for education of the contractor in MT. This expenditure will pay for itself in objectivity achieved. Beyond this, the contractor could also be made available for consultation with the guiding DOD/CIA group during the process of combining requirements with knowledge of the MT field to define a technical program and to monitor that program. Thus the educational process will have a continuing payoff for the DOD/CIA community.

III. SCOPE OF THE REPORT

This report was prepared under the basic premise that a review of the status of the MT field and the attitudes and activities of the sponsors of work in the field would lead, of necessity, to indications of action required to create a coordinated program in MT research and development. The results of the survey showed that this approach was possible and that it would be neither desirable nor necessary to inject opinions other than those of the using and working communities regarding the validity of various requirements for MT or of the different technical approaches to the problem. However, where conflicts of opinion within these communities or uncertainties of knowledge on which these opinions are based became apparent, it was considered in order to raise the questions of substance left unanswered as a result of these conflicts and uncertainties.

Since this study was performed primarily for the DOD, it was restricted to consideration of only those aspects of the problem that impinge on DOD interests and activities. Broader questions with implications of national interest have not been dealt with, but merit mention here to indicate the contextual bounds of the present study. These questions are:

1. the need for dissemination of (translated) foreign scientific and technical information for the use of the scientific and technical community at large, regardless of connection with the DOD;

2. the need for translation of documents, scientific and technical or otherwise, from English into foreign languages for various purposes of national policy;

3. the need for use of linguistics and translation on an ad hoc basis as an instrument in the pursuit of limited war and cold war objectives as they arise in various parts of the world;

4. the desirability and possibilities for international cooperation (even with the Iron Curtain countries) in the linguistics and translation fields as a means for improving communication among nations for any purposes; and

5. the general problem of acquisition, indexing, storage, retrieval and dissemination of information, regardless of the language in which the information appears. (The interaction of this problem with that of MT has been considered, even though the broader problem, per se, has not.)

The following discussion will elaborate on expressed DOD/CIA requirements in the MT field, will describe briefly the current technical state of the field, will indicate what actions are being taken unilaterally by various agencies to further (or modify) MT research and development efforts, and will lead to the suggested course of action for the DOD based on the results of this survey.

In the preparation of this report, the writer has discussed the problem extensively with various persons representing the using and sponsoring community. Technical information sufficient for the purposes of the report has been obtained from these discussions, from talks with a few of those active in the technical field, and from various references alluded to in the text and listed on pages 41-43.

Discussions have been held with the following people:

Mr. Paul Borel	CIA
Mr. Paul Howerton	CIA
Dr. J. Kennedy	AFCIN (USAF)
Col. W. A. Williams	AFCIN (USAF)
Lt. G. Friess	RADC (USAF)
Mr. G. Shiner	RADC (USAF)
Mr. A. Favret	USACSI (US Army)
Mr. J. Kullgren	USACSI (US Army)
Mr. G. M. McClurg	ARO (OCRD, US Army)
Capt. D. Higgins	ONI (USN)
Dr. M. Yovitz	ONR (USN)
Mr. R. See	NSF
Dr. J. Griffith	IBM Corporation
Dr. S. Alexander	NBS
Dr. F. Alt	NBS
Mrs. I. Rhodes	NBS

IV. REQUIREMENTS FOR MT

Although this report must be concerned primarily with DOD needs and actions, it is apparent that the CIA must also be included in this consideration. The Agency's Georgetown program is one of the major ones in the field, and one which many believe will lead to effective outputs in the near future. Certainly the proposed expansion of this program[2] may be expected to interact strongly with any DOD plans in the field. It will therefore be considered that, in effect, the CIA represents a part of the using community with which this report is concerned.

There is no single source within the DOD which specifies translation requirements, even within a single service or agency.* There exist few such specifications related to MT in any case; those which do exist occur in both official and unofficial documents, and both are used in support of work in the field. The case for Machine Translation is generally based on the vast numbers of documents that have to be translated[1][2][3] or on specialized needs of a particular service.[4]**

The official documents deal with translation applications and operations[5]** which include (although not all of these appear in the official documentation):

---

* It should be noted that in addition to its representation on the two committees mentioned on page 3a, the DOD has a representative on the Committee on Exploitation of Foreign Language Documents of the U.S. Intelligence Board (chartered by DCID 2/5).

** References (4) and (5) are included in Appendix A to this report.

a. intelligence

survey and scanning[6][7]

detailed analysis[2][6]

b. provision of technical information to the technical community[1]*

c. special purposes unique to particular service needs

The requirements questions of greatest concern with respect to MT[2][3][4][6] are:

a. the quantity of translation (with an association of time delay between accession of a document and dissemination in translated form);

b. the quality of translation;

c. the languages to be translated; and

d. the cost of translation.

Finally, there are related requirements for abstraction and indexing which impinge on the translation problem.[5][7][8]

While languages can be listed in order of importance or interest for translation at any time, and translation costs are measurable and perhaps predictable, the specification of quantity and quality are exceedingly difficult because both depend heavily on application. There is, furthermore, a tendency to temper the requirement with what is believed can be achieved, both in the areas of output volume and quality.

It will be of value to examine in some detail these various requirements and the positions of the users.

---

* Page 124, Hearings

A. Quantity or Volume of Translation

There are three documents that deal at length with the volume of translation that may be required.(2)(3)(6) All list the amount of published material that might be available for translation - nearly one billion words from the USSR in 1959, for example - and then indicate how much of this material it may be desirable to translate. This is expressed variously:

> "Survey results of the language-translating requirements of the intelligence agencies [not documented] indicate that about 384.4-million words of foreign languages currently need to be translated."(3)

> "If even half of the [Russian] scientific material were worth translating, we would have a total load of over one million words per day for every day of the year . . . [and later] . . . we must set up a center which will be capable of translating approximately one million words per day. . ."(2)

> "Figure 4 shows the millions of words of Russian material which would be useful in translated form to the USAF . . . [the number is 520 million] . . . A 'useful' document is defined as one which the Air Force considers important enough to warrant procuring from Russia."(6) [The criteria for judgment of importance and the analysis leading to the numbers are not given.]

Juxtaposed as they are, these statements show an uncertain knowledge of what is really needed. This uncertainty is not necessarily of grave concern, however. The three references taken together demonstrate clearly that today's manual translation effort falls far short of the needs of the community; that the demand represents some large fraction of the published material in languages of interest; and that the volume of published material as well as the number of languages of interest will continue to grow rapidly. The CIA view is that quantity of translation is of secondary concern, because the demand for output will grow with the capacity for it;[9][10] none of the others in the user community with whom this was discussed dissented from this view.

Thus it would appear that current expressions of volume output requirements should be considered as orders of magnitude only, with precise numerical specifications having secondary importance. Such estimates will undoubtedly have to be made (and will be possible) more precisely later, when expenditures for obtaining the output (whether by MT or manual translation) will become much larger. At this stage, however, with the translating community swamped and an almost complete lack of experience to indicate the volume of usable MT output that can be expected from a facility, the uncertainty appears to be acceptable to the using community.

A related question that appears is whether the vast outputs desired can be of use or can even be handled by the users for whatever purposes.[6][10][11] This question is usually answered by the statement

that not all users will use all the output, but rather that a library will be provided from which various users will draw material of specialized interest (Cf. Ref. (6), page 1).

It should also be noted that all requirements for large volumes of machine translation output include not only the translation itself, but rapid, automatic input and output commensurate with the quantity of translation performed. There are no disagreements on this question.

B. Quality of Translation

No questions are more subject to argument and variations of interpretation than those of the meaning of quality of translation and the quality desired. Some attempts at definition concern themselves with the percentage of words that are translated with correct meaning. These attempts are criticized on the grounds that (a) "percentage of accuracy" cannot reflect the effects of both the numbers of words translated incorrectly and their frequency of occurrence in particular documents, and (b) word translation errors cannot be considered separately from grammatical and syntactical errors.

Most definitions are based on "transference of meaning." But there are various problems associated with this definition, too, varying from the need to include real-world contextual background(12) to language simplification to ease the coding problem in the translation process.(13) Some experiments have been carried out to determine whether translations of particular quality convey meaning to people with various backgrounds in the language and in the field of the

translated material;[2] they demonstrated a need for good background in the field, at least. But this need may be very much relaxed in the case of scanning and screening[7][1]* to identify content for later detailed translation.

It is appropriate to consider here the categorization of translations as "pidgin," "pedestrian" and "poetic."[11] If "poetic" means an elegant translation with great attention to style, such as might be expected in an English rendering of Dr. Zhivago, then none of the potential users in the DOD/CIA community demands "poetic" translation. If "pedestrian" means that meaning is transmitted somehow, without too much difficulty in continuity of reading, then this is what is wanted. But it is quite possible that such results can be obtained from "pidgin" English as well. "Pidgin" English may result from gross and deliberate oversimplification of language to ease the synthesis problem (e.g., subject always precedes predicate; plural of mouse is mouses), or it may result from a syntactical scramble due to inadequacy of the analysis and synthesis rules. Meaning may very well be preserved in the first case but lost in the second. Thus the application of terms such as "pidgin" or "pedestrian" requires a certain amount of analytical insight.

In all discussions of quality, the danger is pointed out that incorrect but smooth translation may in fact mislead the reader

---

* Cf. page 136, Hearings

dangerously. The "incorrectness" may, however, creep in through wrong selection of alternate meanings or through simplifications inherent in the translation rules (e.g., prefix in may be said always to mean not; word flammable appears, meaning burnable; inflammable appears, is translated as not burnable). Elimination of errors requires different procedures in the two cases; the second problem may be more tractable than the first.

Thus the question of quality is intimately related to that of application and also to the method of translation. The point of common agreement is that elegance of language is not needed for the user community we are considering here. Beyond this, it appears that a measure of quality will remain elusive for the foreseeable future, and will start to be attainable only after some considerable experience with MT outputs, and experimentation with their impact on the readers. Specification of requirements for quality, in the face of these difficulties, will not be simple.

C. Languages

Russian is universally considered to be the language of greatest interest and importance for MT application and is the one receiving most current attention.* Chinese and German are given second and third place in order depending on the user;(3)(4)(14)(15) Arabic is also

* Support for efforts to apply MT to languages other than Russian is on a much smaller scale than for Russian, and in the case of Chinese and Arabic, it has been virtually non-existent or is just beginning.

considered to be of great importance. Other languages of interest include Japanese, various Slavic languages, French, Scandinavian, and others that arise from special applications. It should be noted that these language translation requirements, after the first four or five, are not necessarily requirements for MT. Machine applications become desirable only after a certain volume of output in the language for which there is a translation demand has been passed;[8] this threshold value is not defined.

The importance of knowing which languages we want translated lies not only in enabling us to apply better direction to current efforts, but in enabling us to anticipate needs early enough for effective linguistics research and development of machine programming in the languages of interest. While it is clear that the experience gained in programming Russian will be invaluable in showing directions for approach to other languages, and in avoiding many of the errors and blind alleys that have arisen in the initial approaches, it is not clear that programming a new language will necessarily be made vastly simpler thereby.[2][16][17] This is true especially when the language families are different (e.g., Indo-European and Chinese), since some work in the field of linguistics[18] implies interrelated semantic and structural problems that can be resolved only through lengthy research in linguistic theory.

D. Cost

It is expected that ultimately the cost of turning out high volumes of translation by machine will be substantially lower than comparable costs for manual translation.(2)(6) This does not appear to be a factor of immediate concern, however, because even production facilities anticipated for the near future will be considered experimental. They will not necessarily, at first, have the advantages of automatic high-speed machinery in the input and output areas, wherein large costs are incurred. There will be uncertain costs of post-editors and other necessary personnel. The costs of research and development to be charged against the cost of production facilities do not appear to have been explored very thoroughly.

The present attitude towards cost is that it represents but one of many compelling reasons to achieve MT production, so that even if cost per word (or cost by any other measure) is initially higher for MT than for manual translation, this will not be a factor militating against the introduction of automatic facilities.

E. Broader Implications

Every agency concerned with MT requirements views the automatic translation process within the broader context of processing the great amount of information contained in foreign documents. There are thus parallel requirements for screening and abstracting of foreign-language material, closely related to the requirements for translation. The need appears in two forms:

a. rough translation of many documents to indicate subject and context, from which those of special interest may be selected for careful translation; and

b. high quality abstracting by machine (through word frequency count, key sentence selection, etc.), with subsequent translation of all abstracts.

While strictly speaking the problem of screening, abstracting, and indexing is properly related to the information storage and retrieval question and only very indirectly to translation, the implications of the relationship are far-reaching. If the great volume of work to be done is for screening purposes, with only a relatively small number of documents needing complete translation, then the need for MT, or the necessary quality of output, may be much smaller than current requirement statements tend to show, and detailed translation might continue to be done manually. In that case, even though MT research might continue for its useful payoffs in the linguistics and information handling fields,* production facilities might be very different from those which would be required if most documents were to be fully translated. However, the need for full translation of documents, leading to a firm requirement for high-volume production facilities, may remain high in any case. Further, the level of research and development effort that will contribute useful outputs in the broader field is certainly not apparent. Clearly, the impact of the information

---

* Machine programming, coding, and algorithms for language transformation

storage and retrieval problem on MT must be considered, even though the former is not necessarily explored in detail as part of a close look at the latter.

F. Specific Expressions of Requirements

The following expressions of requirements by each of the agencies with whom the question was discussed represent both written requirements, where they exist, and stated ones that form the basis of agency actions in the MT area.

ARMY. The Army desires mechanical translation for three applications:

a. general intelligence analysis;

b. selective translation of technical material to aid the Technical Services' long-range R&D planning efforts (this does not include general dissemination of foreign technical material to the industrial community); and

c. translation, in the field, of intercepted documents.(4) Such translations may be "rough" at Divisional level, and "more detailed" at Army level. They would be performed on some of the Army's FIELDATA family of general purpose computers.

Although item c is the subject of the only documented requirements statement, the opinion was expressed that item a, with b as a corollary, creates the primary need for MT; item c is viewed as something useful to have, but would be accepted as a second-order fallout from

the primary effort. This item carries with it a very difficult language problem, because the unpredictability of location of limited warfare actions may cause any of a large number of languages to become important. The anticipated volume of such translation has not been predicted.

Responsibility for statement of Army MT requirements rests in OCRD.

NAVY. The Navy requirement for translation is stated in the ONI "Missions and Functions" document.(5) The applications divide generally into two areas:

a. translation of documents for intelligence purposes; and

b. translation of documents such as letters, invitations, dictionaries of naval terms, and international agreements and rules (such as Air Traffic Control doctrine), that affect or result from naval operations throughout the world.

The Navy suits its output to the translation capability available. Beyond its own Translation Section in ONI, translations performed in foreign locations and on ships by qualified personnel, and translation performed by qualified linguists under a special effort supported in its reserve program, it relies heavily on the CIA to provide translation service. It believes that as capability for high-volume output grows, its needs will keep pace. The Navy requires reasonably good quality of translation and has found MT outputs available to date unsatisfactory.

Language requirements (after Russian and Chinese) shift with the world situation and the Navy's activities. There are 16 primary languages, mainly from the East; a number of secondary languages for which the urgency of translation is much smaller; and seven languages in which it is desired only to identify subject matter.

ONI supports work in the MT area through technical agencies such as ONR.

AIR FORCE. Although there is no official written Air Force requirement for MT (and apparently not for translation, per se), there has been a large variety of needs expressed:

a. detailed analysis of all manner of foreign documents for intelligence purposes;(1)(7)

b. scanning of documents in rough translation(7) for selection of those having particular interest to be translated in detail;

c. indexing and abstracting, with translation before or after the fact (depending on how meaning is best preserved);(7)

d. translation of scientific and technical material for dissemination to the technical community;(1)*

e. preparation in English translation of Aeronautical Charts and Information from foreign countries.

It is desired that the quality of detailed machine translation be equivalent to that of human translators' output, but this quality remains undefined.(14) No translation would be preferable to poor, misleading

* Cf. page 124, Hearings

translation. Despite this, the rough output available today from the Air Force's main effort (at IBM) is considered by some in the Air Force to have some utility.(1)* The requirement for quantity is dealt with in Reference (6) and in general indicates such high volume that automatic translation by machine is indicated.

Assignment of responsibility for preparation of requirements and furnishing translations is obscure. The intelligence community (AFCIN) would like to have the translation outputs, but does not desire the responsibility for translation.(15) It will support a research effort now leading to MT production later; it will not support a development program or operate a facility.

Despite General Graul's statement regarding dissemination of foreign technical documents to the scientific community,* the procedure for the community to obtain such documents is not spelled out.(7) There are no plans for general dissemination; rather the user must request the documents through his sponsor, the Technical Service. The request may go to ASTIA, but it is not precisely specified where and by what administrative procedure the translation is done. Reference (6) states that ARDC's translation requirement is small.** Thus, although the Air Force supports MT efforts very substantially, the uses and responsibility for this work are not clear.

---

* Cf. page 136, Hearings

** Page 3

Present language needs are Russian and Chinese. German, French and Spanish can currently be handled without MT. Other language needs will appear as the volume of printed material in a language grows.

CIA. The CIA believes that its requirements have been fully expressed in the Congressional hearings and in Reference (2). Basically, these are "fast, accurate translation" of Russian, with other languages in order somewhat as shown in (3). Anything short of high-quality translation will be considered an interim step. It is uncertain whether continued use of post-editors(2) places MT outputs in one category or the other.

There is also an expressed desire for something short of "perfect" translation to be used by the reader somewhat knowledgeable in a language as a "crutch" while reading in the original. This approach will obviate the need to handle graphical and pictorial material in the automatic translation process.

An output of about one million words per day is desired initially(2) from any production facility, although quantity and cost are not primary considerations for reasons given previously. More detailed expressions of requirements specify a number of scientific disciplines (see page 39) and languages of interest in order of importance. These expressions exist in References (1) and (2), which are considered official.

V. STATE OF THE MT FIELD

This discussion is an attempt to provide some over-all view of the various approaches to MT and to obtain some assessment of where the field in general stands. Sharp differences of opinion exist regarding the validity or prospects of some of the approaches. These differences will be brought out, not to demonstrate the correctness or incorrectness of any view, but to show the difficulty of reaching an objective assessment of the technical efforts.

A. The General Problem

The machine translation process can be characterized by the following diagram:(17)

PROCESSING

Input	Analysis	Transfer	Synthesis	Output
original language	1. word meanings 2. syntax 3. semantics 4. real-world references (encyclopedic knowledge of real-world contexts)	1. summary of syntactical analysis 2. representation of rules and structure of both languages	transformation of original to target language with preservation of meaning	target language

The major problems arise in the following areas:

a. programming for dictionary lookup of word meanings, and the integration of such lookup into the broader translation program;

b. stating and coding the (imperfectly understood) grammar and syntax of the original and target language in precise form without ambiguities for the machine program;

c. effecting the transfer from one set of rules to another, to go from one language to another; and

d. selection of the correct meaning, within context, of words or phrases that have multiple meanings (polysemantic).

All of the MT efforts now in progress have grappled with the first problem and are solving it in various ways. The question arises whether it might not be possible to apply a single dictionary lookup technique, by a sort of "SHARE" program, to all the MT programs, and consequently to improve the efficiency of the nationwide operation by reducing parallel efforts. There seem to be two major reasons against this. First, the dictionary lookup process is usually completely integrated with other aspects of the language analysis program; it may have associated with it various codes for grammatical and syntactic identification of words,[19] and a given word may appear with its stem and inflectional endings divided in different ways for purposes of machine coding. Secondly, the machines used for translation, being in general associated with MT as general purpose tools on a research project, have many different memory capabilities and forms of address demanding different approaches to the lookup problem.

All of the efforts that are near yielding machine outputs are grappling with the problems of grammar and syntax.* The difficulties in this area arise from the fact that the rules of grammar and syntax in any language have evolved through usage rather than through any logical formulation. Thus the attempt to establish a precise formulation of language structure must contend with numerous exceptions to rigorously defined relationships among words, phrases, and even thoughts. This problem also underlies the difficulty in effecting transfer from the rules of one language to those of another.

Problems of multiple or ambiguous meaning arise because it is virtually impossible to program into a finite computer the contextual references that determine the choice among various meanings or usages of a word. (This is illustrated by the incident, famous in MT circles, in which "hydraulic ram" was translated from English to Russian and then appeared in English again as "water goat.") Semantic questions are the subject of a small number of very long-range research projects. The more immediate efforts handle the problem by listing various possible meanings for the reader to choose from[1][19] or by choosing a likely meaning according to probability of occurrence associated with surrounding or modifying words in a sentence.[13] Some sources express doubt that

---

* Grammar deals with word structure and inflections of words - word type, gender, tense, number, etc. Syntax deals with the relationships among words in a sentence - placement of subject and predicate, relationships of adverbs to verbs, etc. Neither is concerned with meaning or semantics.

problems of polysemy can ever be solved completely;[12] it is for this reason that the possibility of smoothly reading but incorrect or misleading translation remains a worrisome spectre.

B. Overview of the Field

There are some thirteen contracts of various kinds, sponsored by the CIA, NSF or the services, dealing directly with MT, as well as a number of contracts peripherally related to the translation problem or dealing with hardware for print reading and machine memory storage. The essential characteristics of these efforts are summarized in Appendix B, and a listing by sponsorship and categories of application is given in Table I.*

The approaches to the MT problem are fewer than the number of contracts might imply. The Cambridge and Milan groups are working on semantic problems with a very long-range view, and there are strong overtones of semantics, or rather, general theory of language, in the work of MIT and Hebrew University. The other efforts in this country have all been concerned, in one way or another, with problems of lexical equivalents (dictionary lookup), syntactical analysis and synthesis. The IBM effort to date has been the most frankly concerned with lookup problems exclusively, but is now starting to explore syntactical analysis.[14]

The other approaches may be characterized roughly as "student" and "teacher." In the "student" approach, it is assumed at the outset that

* This information was obtained from References (1), (12), (17) and (20).

nothing is known about the language. A body of rules describing the language is built up from analysis of running text until, it is hoped, a sufficiently large sample has been analyzed so that most (or at least the most common) linguistic situations will have been encountered. The rules thus determined are then used to translate texts newly encountered. In the "teacher" approach, the programmer inserts into the machine all his knowledge about the language, simplified and codified as much as possible to stay within the machine capacity.

Within these broad divisions, various "schools," singly or multiply occupied, have grown. The Georgetown program has considered problems of word-equivalent lookup as well as the questions of analysis, transfer and synthesis, using generally the "student" approach. That approach is applied even to the dictionary problem, since words included in the dictionary are selected from experience with running text in a particular discipline (organic chemistry has been used for initial research). The work at Ramo Wooldridge and Rand may also be considered to use the "student" approach, although prior knowledge of the language is not neglected. Wayne State University is closely connected with the Ramo Wooldridge effort through P. Garvin of Ramo Wooldridge, who acts as a consultant for Wayne.

The most widespread "school" is that flowing out of the work of the National Bureau of Standards, which uses a "teacher" approach. In that program, a sentence is divided into its constituent clauses and phrases;

words in each are examined in order from left to right; from each word the grammatical and syntactic form of the next is predicted; most probable meanings of words are selected; and the entire structure is reviewed through a series of successive iterations until a satisfactory translation (or an impasse) is reached. This general approach is also used by Harvard and A. D. Little and is closely related to the work of RAND and perhaps the University of Pennsylvania.

Finally, the University of Texas, in what may be called a "student" approach, compares manual English translation of German text with the original to search for rules of analysis and synthesis applicable to MT. The work of the University of California at Berkeley, reminiscent of that of the Texas group, also takes a "student" approach by analyzing running texts in a field to establish syntactic and semantic rules.

Now, where do all these efforts stand? The work at MIT, Hebrew University, Cambridge, Milan, and Texas is long-range research and is not expected to produce any results with early applicability to production output. It would appear from the available information that the Harvard, Wayne State University, and Berkeley programs are also in this class. The NBS program is definitely aimed (insofar as the sponsor is concerned) at achieving "production" capability at an early date; since it was one of the late starters (1958), it is not as close to this point as some of the programs that have been going for a longer period.

The IBM, Georgetown, Rand, and Ramo Wooldridge programs are considered to be most nearly ready for tests of quality of output and productive capability.(21) Of these, the Rand work appears to have been diverting in the direction of screening, abstracting, and data retrieval, with translation assuming a secondary aspect;(22) while Air Force support for the Ramo Wooldridge effort is being cut off (as in fact are the Air Force-supported university research projects).* Thus the major contenders for large scale application in the near future are the programs of IBM and Georgetown, and eventually that of NBS.

It is claimed by its RADC sponsors(22) that the IBM program, particularly in anticipated improved versions that pay greater attention to syntactical analysis, offers sufficient capability for meaning transfer to satisfy intelligence needs for screening and scanning, and that this is the primary intelligence application. This view is not concurred in by any of the other members of the DOD/CIA intelligence community (including AFCIN, which, although it believes the IBM approach may have promise, is not necessarily committed to it wholeheartedly). Further, and more important, the possibility for future development of this particular approach to effect better transfer of meaning is viewed with skepticism by the Army and CIA.

The Georgetown approach is likewise held in particular favor by its sponsors, while some other members of the community deprecate its success

---

* See page 38.

and its potential. One of the problems it faces is that, in taking the "student" approach, it may do a fairly successful job of translation when the text includes only linguistic situations that have been encountered previously and accounted for in the program, but it cannot cope with new situations and therefore fails to translate intelligibly. It is therefore necessary to amend the program, which becomes cumbersome and unwieldy. Further arguments arise with respect to the specialization of the Georgetown programs in particular scientific fields, and the implications for the size of dictionary and associated machine storage. While the program aims to have lexicons of some few tens of thousands of entries in each field, it is claimed elsewhere that a lexicon of half a million stems, implying several million words, will be necessary for effective translation of scientific material.

A partial solution to these problems is offered in acceptance of the need always to post-edit the machine output.(2) But this requirement is, also, not necessarily accepted as workable by others in the community. Certainly the Army's effort at NBS, which is said by some to offer the greatest promise of success of all the programs,(1)* contemplates no such necessity.(13) It is not clear, moreover, to what extent the post-editors must be knowledgeable in the scientific field of the translated paper and in the language. This question must be resolved experimentally.(2)

* Conclusions of the Report on the Hearings

Clearly, questions are raised here regarding the acceptability of certain types of output (quality) and regarding technical matters inseparable from the translation programs, which can only be resolved through the detailed explorations of those knowledgeable in the intelligence, scientific and MT technical areas.

Other questions, less difficult, exist on the periphery of the MT problem. These relate to the need for high-speed, automatic input-output devices to keep all parts of the translating system operating at compatible speeds. All parts of the user community are agreed on the need for such devices. The Air Force believes that the Baird-Atomic reader, essentially mechanical in principle, will provide a workable solution compatible with both the Georgetown and IBM translating programs. The Army and CIA, however, believe that the electronic approach to the problem being developed by RCA offers promise of much greater flexibility and scope (number of type fonts without the need for changing them, etc.), and will prefer to await that development, which has now reached the testing stage.

Finally, mention must be made of the question of high-speed output. In general, attention and interest have been centered on immediate problems of translation itself and input to the translator, and most work to date has been satisfied with more or less standard off- or on-line computer output. It has been indicated by the CIA that the high-speed Analex printer, which prints 1000 lines per minute and is an operational piece of equipment, can meet the output requirement.

C. Actions Pending

What is going to be done by the DOD/CIA community about MT in the near future? Pending actions vary from drastic reduction to substantial increase of support for various efforts.

The Army believes that the NBS approach is the most promising of all those in the field and will continue to support that program, as well as the longer-range research program at the University of Texas. Neither program demands a large expenditure; when a working translation program exists it will undoubtedly be connected with the FIELDATA computers, which are supported by many other requirements.(23) The Navy (ONI) expects to include, next year, a half-million dollars earmarked for basic research related to MT in its support of the ONR information systems program.(15)

The greatest perturbations on existing programs will come from the Air Force and CIA. The Air Force (RADC), faced with a lack of funds and low priority for MT, has eliminated funds for all of its university programs (see Table I) and the Ramo Wooldridge program, and will continue to support only the IBM effort.(22) While it does not believe that this effort has progressed enough to warrant acceptance of the IBM proposal for a National Translation Center (nor that the Air Force should support, alone, such a center),(14) it does believe that this program, in its future developments, offers the opportunity to develop

a center to meet Air Force needs.(22)* There is also the question of a substantial prior investment to protect.

The CIA intends to establish a translation center on an experimental basis, built around the Georgetown translation program. By next spring, lexicons for this program will exist in the fields of organic chemistry, economics, geography, physical chemistry, high-energy physics and solid-state physics. A card-punch center has been set up in Germany to prepare inputs for the program; an automatic print reader is intended to become part of the translation complex.

---

* It should be noted here that this action is taken by ARDC as part of its responsibility for equipment development. The AFCIN position, noted earlier, is that it would prefer not to participate in either development or operation of such a center.

TABLE 1

U.S. RESEARCH AND DEVELOPMENT IN MACHINE TRANSLATION AND RELATED FIELDS*

Translation

By Language

Russian-English
Georgetown (2)
Harvard (4)
U. of Calif. (6)
U. of Washington (8)
Wayne State U. (9)
U. of Milan (10)
Indiana U. (12)
Ramo-Wooldridge (14)
Rand (16)
IBM (17)
A. D. Little (18)
NBS (25)

German-English
MIT (1)
U. of Texas (7)
U. of Milan (10)

French-English
IBM (17)
Georgetown (2)

Italian-English
CLRU (3)
(also Latin-Eng.)
U. of Milan (10)

Chinese-English
U. of Milan (10)
U. of Washington (18)
U. of California (6)

English-Chinese, Arabic
Georgetown (2)

English-English
U. of Pennsylvania (5)
PRC (19)

General Language
CLRU (3)
Hebrew U. (11)
Indiana U. (12)
PRC (19)

By Approach

Lexical, with some Syntax
Georgetown (2)
U. of Washington (8)
IBM (17)

Syntactical
MIT (1)
CLRU (3)
Harvard (4)
U. of Pa. (5)
U. of Texas (7)
Wayne State U. (9)
Indiana U. (12)
A. D. Little (18)
NBS (25)

Syntactical, with some Semantics
U. of Washington (8)
Hebrew U. (11)
Ramo-Wooldridge (14)
Rand (16)

Semantical
MIT (1)
U. of Milan (10)
Hebrew U. (11)
PRC (19)

Sponsorship

USAF
U. of Washington
CLRU, Harvard } Through NSF
Indiana
U. of Milan
Syracuse
Ramo-Wooldridge
Baird-Atomic
Rand (OSR)
IBM
PRC
Intell. Mach.

USN
Wayne State U.
Hebrew U.
Hydel, Inc.
Int'l Telemeter
MIT

US Army
NBS
U. of Texas

NSF
MIT
CLRU, Harvard } with USAF
U. of Pa.
U. of California
U. of Washington

CIA
Georgetown

Direct M.T. Contracts → Peripheral → Related

Contracts	Peripheral	Related
U. of Washington	Indiana	Syracuse
CLRU	Milan	Baird-Atomic
Harvard	PRC	Intell. Mach.
Ramo-Wooldridge	U. of Pennsylvania	Hydel
Rand		Int'l Telemeter
IBM		
Wayne State U.		
Hebrew U.		
MIT		
NBS		
Texas		
Georgetown		
U. of California		

Information Storage and Retrieval

Research and Theory

MIT (1)
U. of Pa. (5)
PRC (19)

Equipment Development

Input
Syracuse (13)
Baird-Atomic (15)
Intell. Mach. (22)
Rabinow (23)

Memory and Search
IBM (17)
Hydel (20)
Int'l Telemeter (21)
Rabinow (23)

* Numbers refer to listings in Summary, Appendix B.

REFERENCES

(1) Research on Mechanical Translation

a. Hearings Before the Special Investigating Committee of the Committee on Science and Astronautics, U.S. House of Representatives, Eighty-Sixth Congress, Second Session (May 11, 12, 13, and 16, 1960).

b. Report on the Hearings - House Report No. 2021, Union Calendar No. 895, June 28, 1960.

(2) Howerton, Paul W. The Parameters of an Operational Machine Translation System. Paper read before the National Conference of the American Documentation Institute, Berkeley, California, October 27, 1960.

(3) Planning Research Corporation (Military Systems Research Division). Preliminary Survey of the Need for Language Translation in the United States Government. Report prepared for International Business Machines Corporation, Research Center, Yorktown, New York, April 12, 1960.

(4) Stone, W. W., Jr., Lt.Col., Acting Chief, Research Division, OCRD, US Army. Memorandum for Director of Research; Subject: Machine Translation of Languages. April 2, 1958.

(5) ONI Instruction 05430.2, Statement of Office of Naval Intelligence Missions and Functions, OP-923M4 (P. 87) and OP-92263 (P. 52), June 11, 1959 (Confidential)

(6) Friess, G., Lt. Analysis of Air Force Language Translation Requirements, Report from Rome Air Development Center, Intelligence Laboratory, Language Data Handling Section, June 1960.

(7) Discussion with Dr. J. Kennedy, AFCIN, October 21, 1960.

(8) Discussion with Mr. A. Favret and Mr. J. Kullgren, USACSI, October 18, 1960.

(9) Discussion with Mr. P. Borel, CIA, October 7, 1960.

(10) Discussion with Mr. P. Howerton, CIA, October 14, 1960.

(11) Tyaack, F. H. Mechanical Translation, Institute for Defense Analyses, Research and Engineering Support Division, Report IM-209, June 21, 1960.

(12) Bar-Hillel, Y. Report on the State of Machine Translation in the United States and Great Britain, Hebrew University (Israel), Technical Report T-1, prepared for ONR Information Systems Branch, February 15, 1959.

(13) Mechanical Translation Group, Applied Mathematics Division, A New Approach to the Mechanical Syntactic Analysis of Russian, NBS Report 6595, November 10, 1959.

(14) Discussion with Col. W. Williams, AFCIN, October 27, 1960.

(15) Discussion with Capt. D. Higgins, ONI, October 14, 1960.

(16) Discussion with Dr. F. Alt, NBS, November 3, 1960.

(17) Discussion with Mr. R. See, NSF, October 4, 1960.

(18) Whorf, B. Language, Thought and Reality. Technology Press of MIT and John Wiley & Sons, Publishers, New York, 1956.

(19) Giuliano, V. E., and Oettinger, A. G. Research on Automatic Translation at the Harvard Translation Laboratory. Information Processing - Proceedings of the International Conference on Information Processing, UNESCO, Paris, June 15-20, 1959. (page 163)

(20) National Science Foundation, Office of Science Information Service. Current Research and Development in Scientific Documentation, No. 6. Spring 1960.

(21) Alexander, S. N., NBS. Memorandum for Files; Subject: Summary of Discussion at Meeting of Interagency Committee on Machine Translation Research. October 4, 1960.

(22) Discussion (telcon) with Lt. G. Friess and Mr. G. Shiner, RADC, November 9, 1960.

(23) Discussion with Mr. G. McClurg, ARO, October 27, 1960.

The following references have not been cited in this report, but are of interest and have contributed to the information contained herein:

(24) Yngve, V. H. The COMIT System for Mechanical Translation, Op. Cit. Ref. 19, page 183.

(25) Harper, K. E., and Hays, D. G. The Use of Machines in the Construction of a Grammar and Computer Program for Structural Analysis. Op. Cit. Ref. 19, page 188.

(26) Takahashi, S., Wada, H., Tadenuma, R., and Watanabe, S. English-Japanese Machine Translation. Op. Cit. Ref. 19, page 194.

(27) Booth, A. D., Brandwood, L., and Cleave, J. P. Mechanical Resolution of Linguistic Problems. Academic Press, Inc., Publishers, New York, 1958.

(28) Oettinger, A. G. Automatic Language Translation. Harvard University Press, Cambridge, Massachusetts, 1960.

(29) Bar-Hillel, Y. The Present Status of Automatic Translation of Languages. Advances in Computers, Volume 1, edited by F. Alt; pages 91-163. Academic Press, New York, 1960.

(30) Summary of the Proceedings of the Wayne State University Conference of Federally Sponsored Machine Translation Workers, Princeton, New Jersey, July 18-22, 1960.

a. Read-In. Although photoelectric reading of a printed page of arbitrary format appears possible, this read-in method is probably three to six years off. For read-in during the immediate future, manual typing equipment will be required to digitalize the information.

b. Data Processing and Print-Out. A Mobile Digital Computer (MOBIDIC) is being developed for the Army by Sylvania Electric Products, Inc., to handle data processing for the field army of the future. It is to be van-mounted and delivered by 31 December 1959. The operations of data processing and print-out can be accomplished by specially tailored groupings of components from the MOBIDIC family of data processing equipment.

4. The present expectation is that by 1960 the Army will have a machine language translation capability in the field provided a manual "read-in" device is acceptable and provided machine programs are developed for this purpose. Since each language translated requires a separate machine program, it is necessary to assign a priority to those foreign languages for which the Army desires a mechanical translation capability, either for field use or for translating scientific and technical publications. Arranged in order of priority, these languages are:

a.	Russian	d.	Arabic
b.	German	e.	Japanese
c.	Chinese	f.	Russian satellite countries

5. A program developed for machine translation of a given language depends on the machine to be used. However, several of the steps leading to the program are independent of the machine. Specifically, these are the accumulation and coding of a lexicon and the syntactical and grammatical analyses resulting in syntactical and grammatical rules. Essentially only the programming step depends on the machine. Consequently, any successful mechanical translation effort can be adapted to the MOBIDIC by performing the programming step.

6. Status:

a. The mechanical translation of Russian to English has received the greatest attention. A major effort is being made by the Institute of Languages at Georgetown University under the direction of Dr. Leon Dostert and sponsored by the Central Intelligence Agency and National Science Foundation. The emphasis of this effort is on grammatical analysis with the weakest link being the programming phase. In fact Georgetown University has asked the National Bureau of Standards (NBS) to do their

coding and NBS is presently coding one of the Georgetown approaches. Recently, at the instigation of Lt. Colonel J. A. Ulrich of Diamond Ordnance Fuze Laboratories, NBS submitted to the Office of Ordnance Research (OOR) a proposal for a Russian to English Machine Translation Program. Specifically, this program would be undertaken by Mrs. Ida Rhodes and three assistants. Mrs. Rhodes, a mathematician and linguist, is thoroughly acquainted with the Russian language (Russian born) and also is well known for her skill in programming for electronic digital computers. Although this would be a separate effort at NBS, it would be closely coordinated with the Georgetown University group and would give greater stress to the programming aspects of the problem. It is expected that the NBS proposal will receive favorable consideration by OOR and will result in a contract for approximately $25,000.

b. The Massachusetts Institute of Technology has been making the main effort in the mechanical translation of German to English with Dr. Victor H. Yngve as principal investigator. The University of Texas has submitted a proposal entitled "Proposal for a Feasibility Study of the Machine Translation of German and Russian into English," dated 9 January 1958. This proposal is in four phases with phase I emphasizing language structure and phases II-IV emphasizing machine design and construction. A detailed statement of work and estimated cost ($58,472.60) are included for phase I only. Although Dr. W. P. Lehmann and Dr. C. V. Pollard are proposed as principal investigators, information was received last week to the effect that Dr. Yngve had accepted a position at the University of Texas and desires to continue his mechanical translation work there. Concerning phases II-IV, the need for constructing a special machine for language translation has not been determined. Rather, general purpose computers may be sufficient for this task.

c. Georgetown University, the University of Sorbonne and the University of Algiers have small programs directed at the mechanical translation of French. The mechanical translation of other languages is receiving even less attention.

7. Recommendations: It is recommended that:

a. Present efforts to translate mechanically from Russian to English be followed closely, since the results of any successful effort can be used in the preparation of a program for the MOBIDIC.

b. The Army negotiate a contract with the University of Texas, which would have as its objective, the machine translation of German into English.

c. That an Army-wide project for Machine Translation in the amount of $60,000 be included in the FY 59 budget to support the contract mentioned in b above.

8. Coordination: This memorandum has been informally coordinated with OACSI, ODCSOPS, ODCSLOG, OCOA, Office Chief of Ordnance, Office Chief Signal Officer, OASA, OCAMG, and OCIA.

WILLIAM W. STONE, JR.
Lt. Colonel, U.S. Army
Acting Chief, Research Div.
Office, Chief of R&D

APPENDIX B

SUMMARY OF RESEARCH AND DEVELOPMENT IN MECHANICAL TRANSLATION AND RELATED FIELDS

UNIVERSITIES

1. Massachusetts Institute of Technology (Victor H. Yngve) (NSF) (German-English)

Use of natural language for storage and retrieval in mechanical literature search system. Concentration on sentence structure; breakdown into components. Search for target language equivalents, restructure into target language. Also, work for USN on photomemory.

2. Georgetown University (Leon E. Dostert; A. F. R. Brown; Michael Zarechnak) (CIA) (Russian-English)

Lexical with syntactic rules; attempt to preserve semantics through the rules - use 705, 704 - reorient to 709. Apply to Slavic in general, starting English -- Arabic, Chinese. Syntax rules and words by field of literature. Concentration on chemistry. Simulated linguistic computer: direct coding, use of translation rules - French-English, Russian-English.

3. Cambridge Language Research Unit, Cambridge, England (Margaret Masterman) (NSF) (Language - General, Latin, Italian to English)

General translation program, any language to English (Tests on Latin and Italian). Thesaurus approach (idea-related rather than word equivalents). Program for analyzing sentence structure of arbitrary sentences, from word classes. Punched-card techniques.

4. Harvard University (Anthony G. Oettinger) (NSF) (Russian-English)

Automatic dictionary.

Algorithms for: (a) small contextual neighborhoods, (b) sentence constituents. Automatic classification. Theory of syntax - Russian; artificial languages, e.g., computer. Application of NBS "predictive" techniques to Russian and English.

5. University of Pennsylvania (Zellig Harris, Jr.) (NSF) (English-English)

Application of Linguistic Transformations to information retrieval. Sentence and subsentence structural analysis - computer program (UNIVAC) for recognizing syntactic structure of English sentences. Future, generalization of computer theory, linguistic methods, programming and coding, Program for Transformation Analysis.

(ONR) Information Retrieval Study: Nature of desirable information for retrieval, explicit, to guide design of retrieval system. Automatic indexing and search.

6. University of California, Berkeley (Sydney M. Lamb) (NSF) (Russian-English; Chinese-English)

Russian Language Analysis - lexemes and syntax machine analysis giving alternate translation choices to be made. Build-up of dictionary knowledge incorporated to reduce uncertainty. Coding for operation on 704. Chinese just beginning.

7. University of Texas (W. P. Lehman) (USASRDL) (German-English)

Linguistic analysis, using computers. German-English using "Stochastic Grammar." Concerned with dynamic character of language.

8. University of Washington (Erwin Reifler) (RADC, NSF) (Russian-English; Chinese-English)

Review of status of field just beginning for A.F. MT approach: lexicographical; uses photoscopic disc. Started next phase - logical programs for automatic resolution of some problems in grammar and semantics. Chinese just beginning under NSF support.

9. Wayne State University (Harry H. Josselson; Arvid W. Jacobsen) (ONR) (Russian-English)

Translating Russian Mathematical Texts. Compare manual English translation with Russian and modify former to assume Russian word structure. Machine computation programs for syntactic analysis - IBM 650 and 709.

10. University of Milan, Italy (Silvio Ceccato) (RADC) (Russian-English; some Italian, German, Chinese)

Research into thought processes via operational analysis. Description of correlation sequences, used to describe semantic "bridges" between languages. Programming 704 for 1962; trials on 650. Classification system of meanings in various languages to substitute for experience matrices of original and target language users.

11. Hebrew University, Israel (Yehoshua Bar-Hillel) (ONR) (3 contracts) (General Languages)

Survey of status of mechanical translation research. Analysis of theoretical limits of MT capability, including mechanization of syntax and difficulties caused by polysemy. Studies of theory of literature searching with goal of mechanization. Studies in mathematical linguistics; theory of linguistic models. Mathematical description and correlation among grammars.

12. Indiana University (Thomas A. Sebeok) (NSF) (Russian, USSR Dialects, General Language)

Structural models of language aimed at mechanical translation. First step coding for 650; second step extend program to syntax and metrics.

13. Syracuse University (USAF)

Mechanization of photo handling, in connection with Baird-Atomic print reader.

INDUSTRY

14. Thompson Ramo Wooldridge, Inc., Los Angeles, California (Don R. Swanson; Paul L. Garvin) (USAF-RADC) (Russian-English)

Syntactic analysis assisting dictionary lookup. Use of "fulcrum" (e.g., predicate) to obtain information for analysis of remainder of sentence. Building rules for multiple-meanings.

15. Baird-Atomic, Inc., Cambridge, Massachusetts (John A. Fitzmaurice) (USAF-RADC)

Print reader.

16. Rand Corp., Santa Monica, California (David G. Hays; Hugh Kelly) (USAF) (Russian-English)

Syntactic analysis - sentence structure, grammatical transformations, word derivations. Attempt to bring in semantics. Being machine programmed; "Semi-Automatic." Working on physics text.

17. International Business Machines (Gilbert King) (USAF - RADC) (Russian-English, French-English)

Table-Lookup techniques, intrinsic address. Use of photoscopic disc. Translation is word for word with some overlay of syntax in form of stored word groups or idiomatic phrases.

Also company-run program (Advanced Systems Development Div., Yorktown Heights, New York, T. R. Savage) developing automatic abstracting and indexing techniques. Key word technique with SHARE-704 program.

18. Arthur D. Little, Inc., Cambridge, Massachusetts (Vincent E. Giuliano) (Russian-English)

Interest in production development of experimental MT systems. Working on automatic syntactic analysis with Harvard, using NBS "predictive" technique.

19. Planning Research Corporation, Los Angeles, California (H. P. Edmundson)

Information coding via simplified English language in man-machine system. Retrieval of information which may or may not be transformed in context or reference, through the simplified language.

For USAF-RADC, a program in document analysis. Indexing and abstracting via word frequency and topic sentence frequency count.

For IBM, preliminary requirements analysis of MT. Essentially, count of documents and words translated as compared with those available and those published.

20. Hydel, Inc., Waltham, Massachusetts (USN)

Component Designs - High Speed Photo-memory.

21 International Telemeter Corp., Los Angeles, California (USN)

Development of high capacity photo-memory (photoscopic disc used in IBM MT effort).

22. Intelligent Machines Research Corp., Alexandria, Virginia (USAF)

Work in print readers.

23. Rabinow Engineering Company, Takoma Park, Maryland (Jacob Rabinow)

High capacity information storage devices. Print readers, using optical scanning and matching techniques.

GOVERNMENT AGENCIES

24. National Science Foundation (Burton W. Adkinson; Richard See)

Coordination and sponsorship of various programs in MT. Grants given to MIT, Georgetown U., Cambridge (England), Harvard U., U. of California, U. of Pennsylvania.

25. National Bureau of Standards (US Army, OOR) (Ida Rhodes; Samuel N. Alexander) (Russian-English)

Use of grammatical analysis and predictive approach in MT. Parsing of sentence, prediction of following structure, correlation of prediction with occurrence. Dictionary for word lookup, applied to 704 (Rhodes).

Work on abstract pattern recognition using character recognition with automatic scanning for input and output display. Compression and regeneration of graphic information. (Alexander)

26. US Army (Gregg McClurg, ARO; Lester Geiger, USASRDL; Andrew Favret)

Support of work at NBS and University of Texas.

27. US Navy (Marshall Yovits, ONR)

Support of work at MIT, Wayne State University, Hebrew University, and equipment programs in print reading (Hydel) and machine lookup (Int'l Telemetering).

28. US Air Force (Robert F. Samson, RADC: Jack Kennedy, AFIC)

Support of various efforts in the field, including Cambridge (England), Harvard, University of Washington, Milan, Indiana University, Syracuse University, Ramo Wooldridge, Baird-Atomic, Rand, IBM.

29. CIA (Paul A. Borel; Paul W. Howerton)

Support of work at Georgetown. Cooperates with NSF in furtherance of MT work.