[From: *Advances in Computers* 11, 1971]

Automatic Translation of Languages Since 1960: A Linguist's View

HARRY H. JOSSELSON

Wayne State University
Detroit, Michigan

1. Introductory Remarks

1.1 Some Difficulties of Machine Translation

Many changes have occurred in the field of machine translation since 1960.[1] It has been generally acknowledged, as first pointed out by Bar-Hillel, that fully automatic high-quality translation (FAHQT) is exceedingly difficult, if not impossible, to achieve. The human mind can draw inferences which a machine, even one with an encyclopedic dictionary, cannot do [*1*]. To restate this in terms of linguistics, the science that deals with that aspect of human behavior which manifests itself through language, the primary reason for the unattainability of FAHQT is the very nature of language itself. This fact escaped many early researchers in machine translation, particularly those who were hardware-oriented and who proceeded from the very naive position that since everybody speaks a language, everybody should, theoretically at least, be able to deal with any processes which involve language.

Language was considered just a "bunch of words" and the primary task for early machine translation (MT) was to build machines large enough to hold all the words necessary in the translation process. The fact that words in one language had no counterparts in another and that in some cases one word in one language had to be expressed by a group of words in another (which as linguists know is due to the fact that speakers of a language view the world that surrounds them in terms of the structure of their language)[2] was completely ignored [*2*]. Overlooked also was the fact that two given languages may have completely different structures. Thus there sometimes exists between languages a lack of one-to-one correspondence of what is known as "parts of speech." The way sentences are put together in languages also differs greatly, even among such genetically related languages as French and Italian. And last, the MT researchers were confronted, and still are, by what came to be loosely termed as "semantics"— the fact that words have more than one meaning, and that sometimes groups of words or even whole sentences may have more than one

[1] The present survey of the state-of-the-art in machine translation is based primarily on the author's work in the field, his contacts with researchers, and on the examination of pertinent literature. His research in MT has been supported by the Information Systems Branch of the Office of Naval Research.

[2] Whorf ([2], p. 93) contrasts Hopi, an Indian language of the American Southwest and what he terms Standard Average European (including English, French, and German). He maintains that there are ". . . connections between cultural norms and linguistic patterns," and that ". . . there is a relation between a language and the rest of the culture of the society which uses it."

meaning in a language. Generally speaking, all the translation problems that were not solved by applying rules derived from grammar were lumped together under the heading "semantics," with an indication that since these were difficult theoretical problems they were to be solved later by various devices. These means included the printing out of the several possible solutions of ambiguous text segments to let the reader decide for himself the correct meaning, printing out the ambiguous source language text, and other temporary expedients.

1.2 The Nature of Language and Problems of Language Description

All these difficulties, which of course gave rise to clumsy and what is worse, sometimes outright incorrect translations, were due to the fact that before one can deal successfully with MT, which after all involves language, one must know much more about language. Particularly one must understand the rules under which such a complex system as human language operates and how the mechanism of this operation can be simulated by automatic means, i.e., without any human intervention at all. The first task, linguistic description, is an enormous one that has required and will continue to require a great deal of effort on the part of many serious researchers. Today these efforts are greatly aided by the computer, which assists in formulation of the problem, in testing the proposed solutions, and in proper storage and retrieval of the acquired linguistic information.

The second problem, the simulation of human language behavior by automatic means, is almost impossible to achieve, since language is an open and dynamic system in constant change and because the operation of the system is not yet completely understood. Suppose it were theoretically possible to achieve a complete description of how a given language operates. In the time it would take to compile this description the language would have changed again. It becomes clear, therefore, that the best possible solution for MT, both theoretically and practically, is symbiosis between man and machine.

1.3 MT as Symbiosis between Man and Machine

This man-machine interaction is not to be limited to post-editing of machine-produced texts alone. Rather it is to be viewed broadly as a collaboration between the two in all phases of MT, each partner doing what he can do best most efficiently at a given stage of the translation process, as already suggested by Bar-Hillel (see [*1*] pp. 95-98). Before my views concerning the details of this partnership are spelled out (see Section 6.2), it might be of interest to discuss briefly some of the ideas concerning man-machine interaction in MT advanced by Bar-Hillel in 1960. He states that as soon as one concedes that any system for production of some sort of qualitative MT output will require a collaboration between man and machine, then the “greatest obstacle to practical MT is overcome” ([*1*], p. 95). The obvious corollary to this is, in his words, that, “As soon as the aim of MT is lowered to that of high quality translation by a machine-post-editor partnership, the decisive problem becomes to determine the region of optimality in the continuum of possible divisions of labor” ([*1*], p. 95).

1.4 Practical MT Problems and Some Tentative Solutions

1.4.1 Use of Automatic Print Reader

Bar-Hillel discusses first the necessity for the development of “a reliable and versatile mechanical print reader” ([*1*], p.96) and speculates about the difference in cost between introducing text into the machine by keypunching and automatic print reading. Although admittedly there is no doubt whatsoever that the latter process will eventually be developed and should turn out to be in the long run cheaper than keypunching, it should be pointed out that any cost estimates of any part of the practical MT process are highly conjectural and should certainly not figure in the determination of the practicality of MT at the present time. Any research and development is a, very expensive process and lowering of costs comes only after “the show gets on the road.” It is simply premature even today to justify or dismiss MT for financial reasons alone. It will be shown in Section. 6.2 that practical, quality MT is realizable even today within certain limits. Next, to Bar-Hillel’s prerequisite of an automatic print reader for MT must be added the notion that besides the purely hardware aspect of developing the mechanical reading mechanism there also exists the software aspect—the development of programs necessary for the resolution of some of the text problems for correct input into the computer. For instance, a program must be developed to distinguish between the various uses of the period—end-of-sentence marker, initials indicator, abbreviation, and others. In the case of a print reader for Russian-English MT or the reverse, programs must be written to distinguish for input certain letters common to the Roman and Cyrillic alphabets, like A, B, E, H, O, P, C, X, some of which are identical, like O, and some different, like H, which is N in Russian.

1.4.2 Problem of Dictionary Storage and Retrieval

Bar-Hillel also dwells on the need to compile the necessary dictionaries in a form most suitable for computer storage and retrieval. It now seems that the earlier practice of some research groups of completely ignoring existing dictionaries and of compiling dictionaries from scratch, so to speak, has been superseded by efforts to construct computer-stored bilingual glossaries by first reorganizing the published dictionaries in a format specifically designed for the subsequent MT automatic procedures,[3] such as look-up of words, source language sentence analysis, target language sentence synthesis, and all other information incorporated in what is termed the "grammar code" of each lexical entry. Also, during the translation process, in addition to language information normally contained in the grammars and lexicons, the human translator uses other important information not contained in these two traditional sources of compiled linguistic knowledge, information he carries around "in his head" which enables him to solve many complex problems of translation. All this additional information has to be discovered, and stated in rules storable in the computer. Automatic procedures for discovery of this information as well as for augmenting the quantitative content of computer-stored dictionaries from texts also have been proposed [*3*].

1.4.3 Practical MT: Microgrammars and Microglossaries

It is also now generally recognized that the most efficient way to achieve quality translation and to construct translation rules that will hold true for the overwhelming majority of the material to be translated is to confine oneself to a very narrow segment of technical or scientific literature. The narrower the scope of these materials, the easier it is to construct the microglossaries and grammars for each field and to compile rules for resolution of ambiguities through contextual analysis. In this connection perhaps a comment should be made concerning the form in which source language lexical entries should be stored in the dictionary. Bar-Hillel indicates three possible ways of doing this: (1) the exact letter sequence in which they occur in the text, (2) the so-called canonical form, i.e., infinitive for verbs, nominative singular for nouns, or (3) the canonical stems. In the latter ease a portion of the lexical entry is considered as a stem and stored separately from the inflectional endings, necessitating various routines for splitting the text entry into stem and ending and routines for recombining them later during sentence analysis. Bar-Hillel insists that this question must be solved before mass MT production begins

[3] For example, by 1960, automatic dictionaries based on Russian text had already been compiled by the following MT research groups: Georgetown, Berkeley, RAND. Harvard, Ramo-Wooldridge *et al.* For a. description and list of these computer-stored dictionaries see Josselson [*3*]. For a detailed account of an experiment in reversing bilingual dictionaries by means of computer processing see Josselson [*4*].

([*1*], p. 97) and I am inclined to agree with him. The procedure for storing dictionaries I believe to be most fruitful at the present stage of development of computer technology is given in Sections 3.5.1 and 3.6.1.

2. MT Since I960: Aims and Growth

2.1 Change in Research Aims and Approaches

Perhaps it would be useful now to restate the aims of machine translation research as it is viewed by those who believe, that a reasonable solution to producing a quality machine translation product is possible. Then we will review very briefly the history of MT in the United States, in terms of both the shift of emphasis of the philosophy of approach and the activities of the individual research groups. The primary aim of MT research today is to produce the best possible translation, automated wherever feasible, from one language (the source language) into another (the target language) through the combined efforts of linguists, programmers, and research associates from other related fields. A secondary aim at the present, more an outgrowth of early research than an objective in itself, is to develop as far as possible, a complete description of the way language, and more specifically, individual languages operate. Accumulation of such data is invaluable for subsequent efforts to refine and develop MT output. The acquisition of this linguistic information is of the greatest, interest to other fields in the area of information science, such as automatic abstracting, indexing, and content analysis, as well as to linguists and language teachers.

It can be said, therefore, that the mechanical translation process is envisaged as a joint endeavor by linguists, mathematicians, programmers, computer engineers, and systems engineers to develop an over-all program which utilizes high-speed digital computers and other peripheral automatic equipment, such as printers and eventually hopefully also automatic print-readers, to make possible translation from one language, into another. The general aim of serious researchers today has shifted, as urged by Bar-Hillel among others, from fully automatic translation to achieving a relatively good quality product. This ideally should be as close to human translation as possible, but as is conceded by many, it will never equal human translation. This translation product, will be, as it is to a certain degree even today, useful to those who are interested primarily in acquiring information from the source language, not in a never-to-be-achieved elegance of translation. An exceedingly practical subsidiary aim of machine translation is to produce an output which ideally will not exhibit at least the same degree of errors in interpreting source language as committed by a human translator.[4] Thus, to repeat once more, the machine translation process should be viewed as a process in which automatic equipment plays a major role in the symbiosis between man and machine.

In reviewing the history of machine translation in the United States, one should probably start in 1946, when Warren Weaver and A. D. Booth first discussed the possibility of machine translation. A fuller discussion of the early history of MT appears in Booth and Locke [*5*]. In the early days machine translation was viewed as another illustration of the "intellectual capability" of computers [*6*].[5] Computer technology was beginning to develop successfully and there was a desire to prove that the machine was capable of solving all the problems that man can solve. However, the first experiments in machine translation, starting with the trial run in 1953 by Leon Dostert, Director of Georgetown University Institute of Languages and Linguistics, really did not prove anything, particularly as far as the possibility of FAHQT is concerned. This was due to the fact that the first MT experiments were carried out on very limited small texts, with bilingual glossaries and grammars specially tailored for these texts, thus in effect creating an ideal, closed linguistic system in contrast to the openness and dynamics of natural language. The computer programs specifically designed for these small texts, of course, guaranteed the success of these experiments. As a consequence early MT researchers arrived at the conclusion that all that was necessary to achieve practical results in MT was to increase the size of the dictionary and to expand the grammar.

Consequently, the majority of efforts of the MT research groups in the United States and elsewhere were limited to such tasks as construction of schemes for machine-stored bilingual dictionaries; development of grammar codes[6] for dictionary entries, syntactic analysis, including automatic sentence parsing routines; and automatic language

[4] As is well known, the three components of human translation are (1) knowledge of the target language, (2) knowledge of the source language, (3) expertise in the subject, matter to be translated. Those are considered of equal importance in the translation process and a deficiency in any of these can impair the quality of translation.

[5] Similar views were also expressed by Yu. A. Shreider in a report to the meeting entitled "Automatic Translation: Illusions and Reality," cited in [*6*]. I am greatly indebted to the discussion in [*6*] in my further discussion of the state of early MT research.

[6] In MT parlance, the term "grammar code" refers to all morphological, syntactic, semantic, and other linguistic information stored in the computer for each dictionary entry which is necessary for further MT processing.

recognition routines in general, routines for contextual resolution of text ambiguities and other information thought necessary for the development of automatic procedures adequate to handle the translation process. By 1962, according to the National Science Foundation, twelve research groups were actively engaged in these tasks in the United States. The same was also true for machine translation groups in Great Britain, France, the Soviet Union, and on a smaller scale in Italy, Germany, Japan, and other countries.[7] The total number of MT groups in the world reached 48 that year [*7*].

2.2 Federally Supported Cooperation among MT Groups in the United States

Only a few short years after the first organized research efforts in MT were initiated, research personnel and techniques had greatly expanded. Several pioneer researchers, in collaboration with their Federal sponsors (notably, The U.S. Office of Naval Research, The National Science Foundation, and the U.S. Air Force), began to feel the need for cooperative exchange of information among the various research groups to minimize duplication of certain tasks and to attempt to develop MT along more efficient and productive lines. The idea for arranging such an exchange of views about, the immediate problems confronting MT researchers took more definite shape in the course of informal discussions among some of the participants at the National Symposium on Machine Translation" held at the University of California at Los Angeles in 1960 and led to a series of machine translation conferences organized by Wayne State University [8].

2.1.1 First MT Conference at Princeton

The first conference of federally sponsored machine translation groups was conceived of as a working conference to encourage open discussion and cooperation among participating groups. This spirit of mutual exchange prevailed and established the pattern for what became a series of highly successful “Princeton-type” meetings, as

[7] The activities of all research groups, in the United States as well as abroad, conducting investigations of MT-related problems (encompassing software, hardware, and linguistic theory, description and analysis) were documented in a. series published by the National Science Foundation's Office of Science Information Service between 1957 and 1966. This series, entitled *Current Research and Development) in Scientific Documentation,* also includes descriptions of research activities in areas of related interest, e.g., language analysis, systems design, pattern recognition, and information storage and retrieval.

[8] Summaries of the proceedings of these meetings were prepared, printed and distributed to the participants. Copies of these summaries are on file at the Slavic Department of Wayne State University, Detroit, Michigan.

they came to be called. The first meeting was held in Princeton in 1960 and included 26 official participants and several invited visitors, with representatives from MT research groups in the United States, England, and Italy. While the general theme of this conference was dictionary design and the general problem of MT grammar, discussions included such topics as programming strategies, and possibilities for the interconvertibility of the materials, codes, and formats of the various groups. Appraisals were also made of the status of text analysis, dictionary compilation, and grammar coding.

2.2.2 Second MT Conference, Georgetown: Grammar Coding

The second conference was convened at Georgetown University in 1961 and was devoted primarily to problems of grammar coding and the optimal content of the grammar codes of Russian English automatic dictionary formats. Two representatives from each of the following MT groups came to this meeting: Berkeley, Georgetown, Harvard, M.I.T., National Bureau of Standards, Ramo-Wooldridge, RAND, and Wayne State University. The specific aims of this meeting were (1) to discuss codes, coding procedures, and approaches of Russian - English MT groups in the United States; *(2)* to consider the prospects for the automatic conversion of coded materials among the various groups; and *(3)* to decide what material was to be coded and thought to be appropriate for entry into the code. Determination was made of the extant and proposed dictionary coding. Special investigations of Russian grammar by the various groups were noted and listed.

2.2.3 Third MT Conference, Princeton: Syntactic Analysis

The third mooting was held at Princeton in 1962 and concentrated on MT oriented syntactic analysis. The meeting was held to enable the groups to examine together the main problems that had come to light in syntactic, analysis, to compare solutions, and in general to clarify, interpret, and compare the results of various individual endeavors. The discussions included consideration of syntax problems not only in Russian, but in such other languages as Arabic, Chinese, and Japanese. One entire session was devoted to discussion of theoretical models used in syntactic analysis. MT groups, not previously represented, present at this conference were IBM, Ohio State University, and University of Pennsylvania.

2.2.4 Fourth MT Conference, Las Vegas: Semantic Analysis

This meeting held in 1965 dealt exclusively with computer-related semantic analysis. It included a keynote address by the President of the Association for Computational Linguistics, Winfred Lehmann of the University of Texas, and thirteen papers. Besides discussions immediately following the presentation of the above papers, two sessions were scheduled for informal discussion. Six foreign scholars participated: three from the United Kingdom, and one each from Hungary, Israel, and Italy. In addition to observers, eleven federally sponsored groups working in MT and related areas were represented, Also in attendance were representatives of several interested United States Government agencies.

It was generally felt that these meetings benefited all participants in terms of the climate of cooperation created, the informal exchange of views, and the precise definition of problems to be resolved. Various independent scholars and key representatives of research groups were able to examine together the problems of either a linguistic or computational nature that have confronted serious researchers. Also, conference participants were able to compare solutions, question one another on particular points, and in general, pool the results of individual endeavors.

2.3 International Conferences

In addition to the above meetings, two simultaneous developments growing out of the same climate of cooperation should be mentioned. These are (1) the organisation of international meetings on automatic language analysis and computational linguistics, and (2) the establishment of a professional society called the Association for Machine Translation and Computational Linguistics (known since 1968 as the Association for Computational Linguistics).

As to the first development, the following international meetings were held and included papers and discussions on machine translation, among other topics.

(1) First International Conference on Machine Translation of Languages and Applied Language Analysis, held in September 1961 at the National Physical Laboratory in Teddington, Middlesex, England. The papers presented at this conference were published in two volumes *(Proc. 1st Intern. Conf. Machine Translation of Languages and Applied Language Analysis,* H.M. Stationery Office, London, 1962).

(2) 1965 International Conference on Computational Linguistics held in New York. The papers presented at this meeting were made available in the form of preprints only. [9]

[9] These preprints are available from the Slavic Department at Wayne State Univerity, which functioned as secretariat for this meeting.

(3) Second International Conference on Automatic Language Processing held in August 1967 at Grenoble, France. The papers from this conference were bound in one volume.[10]

(4) The 1969 international Congress on Computational Linguistics held in September 1969, Sanga-Saby, in Sweden.

2.4 Formation of a Learned Society

An ad hoc Committee was set up in 1960 to create a professional society to be called the Association for Machine Translation and Computational Linguistics (AMTCL). The members of this committee were participants in research groups working in MT and related areas. The Committee met at Teddington in 1961 at the First International Conference on Machine Translation and Applied Language Analysis.[11] It was moved and seconded at this meeting that the proposed society be formed and the proposed constitution be adopted. Signing of the constitution as well as the balloting and installation of officers took place at the Third Princeton (Syntax) Conference in 1962. Later the same year, the Executive Committee met at the Massachusetts Institute of Technology and planned the first annual meeting which was held in Denver, Colorado, in 1963. Since that time, annual meetings have been held and membership has grown to approximately 650 at the date of present writing. To reflect the interest of its membership, the location of the annual meetings is alternated between the sites of the meetings of the Linguistic Society of America and the Association of Computing Machinery, being held either immediately preceding or following the sessions of these societies. As already indicated above, the name of the society was changed to Association for Computational Linguistics at the sixth annual meeting held at Urbana, Illinois in July 1968. The Association has two publications *MT (Mechanical Translation)* and *tFS* (*the Finite String)*.[12]

[10] The papers presented at this meeting were bound together under the title *2ème conference internationale sur le traitement automatique des langues* and were distributed to conference participants.

[11] The following charter members attended this session: H.Brownson, L.Dostert, H.Edmundson, P.Garvin, D.Hays, K.Harper, H.Josselson, S.Lamb, W.Lehmann, A. Oettinger, E.Reifler, R.See, and V.Yngve.

[12] Information concerning these publications can be obtained from: *Mechanical Translation* (Victor Yngve, Ed.) Graduate Library School, University of Chicago, Chicago, Illinois, 60637; and the Finite String, (A. Hood Roberts, Ed.) Center for Applied Linguistics, 1717 Massachusetts Ave. N.W., Washington, D.C. 20036.

3. Linguistic Problems in MT Research

3.1 Inadequacy of Early Linguistic Models

It was during early MT investigations that most of the researchers became aware, as was predicted by serious-minded linguists, how little was known about the structure and usage of the language they were working with- even about Western European languages, though there has accumulated over the last two centuries a vast amount of detailed description. All this accumulated linguistic information, it turned out, was incomplete, and it became clear that the theories underlying those machine translation efforts fell far short of both the goal of obtaining satisfactory practical results and the attempt to solve the problem of providing adequate models for the operation of languages. The first MT procedures were obviously of an extrapolation type. They were constructed, as indicated above, on the basis of the examination of a certain number of texts and the resulting rules were then transferred (or extrapolated) to other texts. These were clearly *ad hoc* rules, not based on a thorough investigation of the structure of language (see p. 28 of [*6*]).

3.1.1 Quest for Operational Linguistic Models

The emphasis of MT research, therefore, shifted in the mid-sixties to the second, far more important aim, namely, finding out more about how language operates and creating language models. For a brief discussion of contemporary linguistic theories see among others [*9*]; for models see [*10*]. Just as prior to launching a satellite successfully into outer space a mathematical model of the dynamics of the path of the rocket must be constructed, or before measures for controlling the path of the economy of a nation can be instituted a model of the economic activity of a country must be developed, it is necessary to have an adequate model of language before a quality MT product can be obtained. The fact that language is essentially a biologically based system adds to the difficulty of developing language models, and compounds the complexity of the nature of the translation process. Written texts are an output of the operation of linguistic processes governed by language structures. Therefore, every model of language must reflect the operation of linguistic structure. The fact that the actual processes governing language behavior are not well known explains both the difficulty of constructing adequate language models and the necessity of pursuing serious efforts in this direction (see. p. 28 of [*6*]).

3.2 Some Linguistic Problems Facing MT

Some of the specifics of the linguistic problems which have confronted the MT research groups in the United States will now be discussed along with a general overview of attempts to solve them, touching both on theory as well as on practical solutions. Most of the illustrations will be cited from the area of Russian English MT. since this is the area in which the author and the majority of MT groups in the United States have been working.[13]

3.2.1 Area and Units of MT Analysis

Most of the machine translation groups in the United States and elsewhere confined themselves to the translation of materials in natural science physics, chemistry, biology, and mathematics; social science, such as economics; and occasionally newspapers, such as articles in the Russian language newspaper *Pravda.* Generally speaking, the translation unit was the sentence, which in the course of analysis was further subdivided into phrases, words, and morphemes. Automatic analysis of units larger than a sentence were envisaged, e.g., a paragraph, chapter, or book, but up to the present, machine translation research, in the United States at least, has operated within the limits of a sentence. This imposed some limitations upon the resolution of certain morphological ambiguities. For instance, the Russian possessive pronoun его means "his" or "its," depending upon the antecedent. When the latter is not present in the sentence which is being translated, sentences preceding the one containing его have to be scanned for the possible resolution of this ambiguity.

3.2.2 Contrastive Grammatical Analysis

The differences in grammatical structure of the input and output languages also contributed to the difficulties in machine translation. In fact, major efforts of linguists were directed to this problem, both on the analysis and the synthesis levels. Essentially, the linguist in machine translation has had to construct a scheme of contrastive analysis between the source and target languages. It has been pointed out by many linguists that the analysis of a language from the point of view of that language alone will produce one type of scheme. But when one language is analyzed side by side with another, that is, when

[13] The ensuing discussion closely parallels the materials contained in my paper "Linguistic basis of mechanical translation: contributions of standard linguistic theory." presented at the Seminar on Mechanical Translation, held under the auspices of the U.S. Japan Committee on Scientific Cooperation at Tokyo, Japan, April 20-28, 1964.

one language is mapped into another, a different analytical scheme is found to be more useful. In machine translation, contrastive analysis yields the best results. The source language is analyzed not only in terms of itself, but also with a view towards expressing its structure in terms of the target language.

3.2.3 The Lack of One-to-One Correspondence between Languages

As anyone who has engaged in translation from one language to another knows, one of the difficulties in translating, generally speaking, is the frequent absence of one-to-one correspondence between languages on the lexical, morphological, syntactic, and semantic levels. For instance, the Russian adjective used as a noun, зрячий has no exact one-word equivalent in English and has to be translated by a phrase, such as "”one who sees.” Russian lacks the class of article, which English has. Consequently, in translating from Russian into English, one faces the problem of inserting into English, *a, an, the,* wherever these are needed in the English text. In Russian the conjunction пока in the meaning “until” is always followed by the negative не while in English the equivalent construction is affirmative: сиди здесь, пока я не приду –“'Sit here until I come.” The Russian verb бренчать has the general meaning “to tingle.” But the sentence деньги бренчат в кармане is translated into English as “Money is jingling in the pocket,” not “tingling.” Coins “tinkle” in Russian just as a piano “tinkles,” but in English coins “jingle” and a piano “tinkles.” The difficulties that all the above absences of one-to-one correspondence will cause in mechanical translation from Russian into English are quite apparent from the above examples and these problems must be dealt with, if good translation is to be obtained.

Let me cite a few examples indicating how machine translation linguists have attempted to handle the lack of class equivalents between two languages. Frequently the machine is instructed to search the context of a lexical item in the source language in order to render a suitable translation in the target language. Whole classes missing in the source language can also be suppressed, at least partially, in the target language. Lexical items belonging to certain classes can be inserted into the target language as translations for morphemes present in the source language, but missing in the target language. In English, prepositions will sometimes have to be inserted for certain Russian case endings, “of” for the genitive case ending, “to” for the dative, and so on. Words which are followed by negation in Russian, and their equivalents which appear in affirmative English constructions, are appropriately marked in the source language, Russian, accompanied by appropriate instruction to the computer for handling this situation in the target language, English.

Ideally, any translation, human or mechanical, should transfer the content of any utterance, from one language into another. Roman Jakobson expresses this notion succinctly when he says, "Most frequently, translation from one language into another substitutes messages in one language, not for separate code units, but for entire messages in some other language" (see [*11*]). At the present stage of development of machine translation, this ideal state has not quite arrived yet, because of the difficulties of establishing interlingual class equivalents by mechanical means. Yuen Ren Chao, in a paper delivered at the Ninth International Congress of Linguists, pointed out a temporary solution for this problem. ". . . Certain cases of non-correspondence, or at least complicated patterns of correspondence will, at the present state of the science, have to be left to non-machine translation [i.e. human post-editor], I have in mind such cases where one language has one form of structure, say Subject-Verb-Object, and the other language has a similar structure for certain instances, but a different structure for other instances, conditioned by non-structural, but lexical factors" [*12*].

3.2.4 Nested Structures

Another problem encountered in machine translation is nesting, a phenomenon which may occur in any language that grammatically allows sentence structures containing discontinuous constituents. Nesting has been defined as the interruption of a main clause by an embedded phrase or clause, followed by the resumption of the main clause. As has been pointed out by Murray Sherry, of the Harvard Machine Translation group, ". . . a *level* of nesting or a *depth* of nesting can be assigned to every phrase and clause, in a sentence" [*13*]. Hence, deeper levels of nesting may be identified when the embedded phrase or clause itself contains an embedded structure. For example, consider Sherry's illustration in English: "The man who came to dinner ate heartily." On the first level one finds—"The man ate heartily," "who came to dinner" is on the second level, and "to dinner" is on the third and deepest level. If one accepts the above definition of nesting, then the nested structure of the following Russian sentence may be identified: Система, параболическая в смысле И.Г.Петривского, удовлетворяет условию. – "The system parabolic in I. G. Petrovsky's sense satisfies the condition." Here, the main clause система удовлетворяет условию – "the system satisfies the condition" has been interrupted by a modifier phrase.[14]

[14] The nesting concept, the depth hypothesis, and schemes related to the identification of embedded structures have also been discussed by the following: Alt [*14*], Yngve [*15*], and Sager [*16*].

An interesting discussion of major types of embedding constructions in Japanese is made by Charles J. Fillmore, of the Ohio State group [*17*]. He points out that one of the main characteristics of Japanese syntax is the embedding transformation analogous to the English relative clause embedding rule, which attaches sentences to nouns. Since in Japanese the relative clause precedes the noun it modifies, and since the transformation may be repeatedly applied in the generation of a single sentence, multilayered Japanese sentences appear to have a constituent structure which is not altogether compatible with the depth hypothesis. To illustrate his argument, Fillmore, analyzed a Japanese sentence which appeared to exhibit a considerable amount of left branching, as opposed to English which branches to the right.

It is, therefore, quite clear from the foregoing that a significant share of the effort of the American linguist working on machine translation has been devoted to the analysis of the grammatical structures of the languages involved; I shall now discuss the models underlying the analysis as well as the specific approaches and procedures employed by various MT groups both in the United States and abroad.

3.3 Linguistic Theories Underlying Automatic Syntactic Recognition

Following the compilation of grammar-coded, automatic dictionaries, automatic syntactic, recognition and analysis became the primary objective which MT research groups shared in common. Researchers concluded from the beginning, either intuitively or empirically, that the development of syntactic recognition algorithms could only proceed from the establishment of a specific approach to grammar, i.e., within the framework of arriving at a well-defined system of language structure and organization. Moreover, the needs of MT made unprecedented demands for a grammar of precisely formulated specifications.

3.3.1 Theoretical language Models for Syntactic Analysis

The following approaches to grammar have been adopted by practically all MT research groups since at least 1960. The grammatical models themselves may be distinguished according to traditional linguistic criteria as either formal (theories) or descriptive (approaches) in nature. Garvin [*9*] prefers to label the former strongly model-oriented grammars, and the latter weakly model-oriented grammars. In essence, the four major theoretical language models are dependency theory (see. e.g.[*19*]), stratificational theory (see, e.g.[*20*]), transformational theory (see, e.g. [*21,23*]), and formational theory (see, e.g. [*22, 23*].

Dependency theory, which has primarily been employed in the MT activities of the RAND Corporation and Georgetown University, was elaborated by David Hays, and is in principle based on the assumption that words in a sentence are mutually dependent in a hierarchical way.

Stratification theory, which was applied to the MT experiment of the Berkeley research group, is predicated on a concept of language, of which the structure may be said to consist of several hierarchical levels or strata, the lowest of which is sound, and the highest is meaning.

Transformational theory, notably articulated by Noam Chomsky among others, is the basic approach of such MT research groups as the Center for the Study of Automatic Translation at Grenoble, France. and the University of Montreal. This theory proceeds from the belief that a language has certain basic types of sentences called kernel sentences from which the remaining, more complex sentences of the language may be derived by application of transformation rules. A language, therefore, consists of kernel sentences and their transforms.

Formational theory, applied in MT experiments at the Linguistic Research Center at the University of Texas, appears to be closely related to transformational theory. It is based on the assumption that it is possible to formulate a mathematical theory of the formation of symbol strings. Also involved is the necessary creation of meta-languages for processing object languages.

3.3.2 Grammatical Approaches to Syntactic Analysis

Some of the better known grammatical approaches to syntactic analysis are predictive analysis,[15] immediate constituent analysis [*9*], the fulcrum approach,[16] clump theory,[17] and the correlational grammar (see [*2*]).

Predictive analysis was introduced by Ida Rhodes of the National Bureau of Standards MT group, and further elaborated by the Harvard University group headed by Anthony Oettinger. It is based on the assumption that a Russian sentence can be scanned from left to right, and the syntactic role of a given word in the sentence can be determined on the basis of predictions made during the analysis about the word on its left. The hierarchical function of the word on the right is also predicted based on the word being examined.

[15] *Predictive analysis.* various reports of the MT research groups at the National Bureau of Standards and Harvard University's Computation Laboratory.

[16] *Fulcrum,* various reports of the MT group at the Bunker-Ramo Corporation.

[17] *Clump theory,* various reports of the MT group at the Cambridge Language Research Unit in England.

The immediate constituent analysis is based on the view that, taking the sentence as the largest linguistic unit, by a series of consecutive sectionings and subsectionings such an analysis can yield increasingly smaller sentence constituents. This procedure entails multiple, back and forth, scans of the sentence, particularly when automatized.

The fulcrum approach, in opposition to the immediate constituent theory, commences analysis with the smallest constituent of the sentence and proceeds to group constituents into increasingly larger units. This approach, developed by Paul Garvin, has been the basis for syntactic analysis at the Bunker-Ramo Corporation, and a modified version of it has been used in MT research at Wayne State University.

Clump theory was formulated and used at the Cambridge Language Research Unit by A. F. Parker-Rhodes and Roger Needham. This technique for syntactic analysis performs multiple scanning of a sentence and, on the basis of dependency and government relationships, groups related sentence constituents.

Correlational grammar was first developed by Silvio Ceccato in his MT experimentation at Milan in what came to be known as the Italian Operational School. The approach in this system is based on relations created between individual lexical items and their constructed constituents. Basic to this theory is the creation of a table of finite relations that the human mind sets up between the individual items in a train of thought. There is an attempt to correlate human thought with verbal expression.

3.4 Differences between Traditional and MT Grammar

The question that arises quite naturally at this juncture is how machine translation grammar differs from traditional grammar. The differences are conditioned by the fact that the analysis is performed by the computer and not by the human being. The latter has recourse to bilingual dictionaries and written grammars, as well as all the information about language in his head, which he has accumulated during numerous previous translation activities. For machine translation this total information from all possible sources has to be compiled and arranged in a form suitable for processing by the computer. No ambiguities in description will be tolerated by the computer. Every rule must be precisely formulated: all exceptions to rules must be explicitly specified; no “et ceteras” can be allowed, since the computer will not be able to understand them. Linguists traditionally described, but never attempted to test the precision of their grammars. Machine translation has provided the first real challenge to test the power of grammars constructed by the linguists in the course of their research.

3.5 Components of an MT System

The principal steps in the, MT process, based primarily on the procedure employed by the Wayne State University group will now be discussed. For the most part, however, these procedures, with a few variations, were in general use by other MT groups as well.

3.5.1 MT Dictionary Compilation

Generally speaking, the first step in machine translation involves the compilation of a bilingual glossary containing all the lexical, morphological, and syntactic information about each entry, and eventually also semantic information necessary in the translation process. To reduce the complexity of some of the problems encountered in the course of this compilation, these dictionaries dealt usually with a specific science subfield—chemistry, physics, mathematics—the so-called microglossaries. Dictionary information was first punched on IBM cards, and then stored on magnetic tape or in core memory, or both, for dictionary look-up by the computer. Words of the incoming language text to be translated were also keypunched, put onto magnetic tape and then by means of appropriate programs looked up. Lexical items found to be missing during the dictionary look-up process were either transliterated, or computer routines were developed to generate, by speculating on their location in the sentence and on their morphemes, the grammatical information of the missing forms. These forms were appropriately coded by human translator and then inserted into the existing dictionary tape. Using all these devices an updated dictionary tape, was produced. This text tape with the looked-up information was then processed by syntactic programs especially developed for this purpose.

3.5 2 Syntactic Analysis

These syntactic programs further analyzed the input language sentences by employing various parsing routines, which depend to a large extent on the linguistic theory utilized by the MT groups discussed above. The main purpose of these procedures was, of course, to establish the hierarchical relation between these sentence units, viz., subject, predicate, object, and other components.

3.5.3 Target Language Synthesis

After the components of the source language have been identified and appropriately tagged, the target language equivalents of the source language lexical items are printed out in target language order.

Next, appropriate target language items missing in the source language are inserted; in the case of Russian-to-English translation, these items are articles, prepositions, various forms of the verb "to be," "to do," and other function words. Some source language lexical items have to be deleted in the target language. As an example of deletions in the case of Russian-to-English translation, Russian emphatic particles, же, да, and others can be cited. The lack of structural class equivalents between the two languages has also to be dealt with. Resolution of homography has also to be attempted. And finally, the source language items have to be rearranged in target language order.

3.5.4 Treatment of Idioms and Lexical Units

One way of dealing with the lack of one-to-one correspondence referred to above, between languages on the lexical level is to establish a class of lexical units called idioms. Groups of words in the source language have as their lexical equivalent in the target language one or more words. An idiom can also be defined as a group of words the lexical meaning of which is different from the lexical meanings of the words constituting it. Thus, the Russian idiom, несмотря на, the literal translation of which is "not looking at," is best translated into English by "in spite of."

The way idioms were handled by machine translation groups in the United States is related to the dictionary look-up schemes used by them. Some groups, among them Georgetown, RAND Corporation, and Bunker-Ramo, stored idioms in a separate list or table and looked them up prior to the main dictionary look-up operation. The Wayne State University group stored idioms as ordinary dictionary entries, and looked up incoming text in text order. The idioms were identified according to the so-called "longest match scheme," in which a check was first made to determine whether all the constituent parts of an idiom were present. In other schemes, the incoming source language text was first sorted out, either alphabetically or according to some other scheme, e.g., an arrangement of words in an order of descending frequency of use. Idioms were looked up in a special table, as pointed out above, and then the rest of the lexical units of the incoming text were identified.

In glossaries which consisted of full words and stems plus affixes, the full words were looked up first by a matching procedure. For words not found by direct look up, the word was first split into a stem and affixes according to special routines devised for this purpose. An attempt was then made to find a matching stem by checking with a special table of stems. The endings were then also matched against a special table of possible inflectional endings. If a perfect match was obtained in both instances, the full word was again reconstituted and the stored grammatical information attached to it. If a complete match was not obtained, the source language lexical items found were inserted in the eventual printout according to some prearranged scheme, indicated above, i.e., either transliterated, their grammatical information generated, or looked up by a. human being.

After the lexical units of the source language, be they fu1l words, stem, affixes or idioms, have been identified, information needed for further translation processing by the computer, which is usually termed "grammar code" by some machine translation groups, has to be attached to them. The way this information was arranged was governed to a major extent by the linguistic theory employed in the grammatical analysis of the sentence (since, as pointed out above, the translation unit of most machine translation groups has been the sentence), as well as by computer considerations. Here one must distinguish carefully between the linguistic theories which are used by the various machine translation groups, either in expressed form or in a form implied in their statements, and the approaches and procedures used by them in the course of analysis and subsequent computer processing.

3.6 Wayne State University MT Procedures

3.6.1 Wayne State University Grammar Code

The lexical units which comprise the Wayne State Machine translation dictionary are coded with respect to their function in the sentence. Although morphological information is used, the basic criteria for the classification employed are syntactic. Nine form classes are established in the Wayne State University grammar code: (1) *nominals,* which comprise all those words that function as nouns (included here are also personal pronouns and the pronouns кто "who" and что "what"); (2) *modifiers,* which include all words that can modify a nominal (adjectives, certain pronouns, e.g., possessive and demonstrative), numerals and verbal participles; (3) *predicatives,* which consist of all words that can be used as predicates, including finite verbs, short form adjectives, comparatives of adverbs used predicatively, words denoting the so-called "category of state" in Russian — категория состояния — жаль — "it is a pity," можно — "one can," надо — "it is necessary" —; (4) *gerunds;* (5) *verbal infinitives;*[18] (6) *adverbs* and *particles,* which are grouped together for the sake of convenience

[18] Gerunds and infinitives are separated out from the rest of the verb because they differ from the latter in the manner in which they function in the Russian sentence.

since relatively little is done with them in the course of sentence structure determination; (7) *conjunctions,* (8) *prepositions* (these are treated the same as in traditional grammar); and (9) *declined relatives* which consist of relative pronouns like какой, который. — "which," чей — "whose" and which are separated from the other pronouns because of their property of serving as part of an initial clause boundary marker.[19]

The grammatical information thought to be necessary for automatic translation is first entered by coders on appropriate coding sheets and later keypunched onto IBM cards. In the Wayne State University scheme there are two IBM cards for each dictionary entry, one containing the Russian word and the accompanying grammatical information, the other the English equivalent or equivalents of the Russian word. The IBM cards are subsequently stored on magnetic tape, as already indicated.

The first 24 columns of the Russian dictionary card are reserved for the Russian form. Column 25 contains an indication of the form class to which the Russian word belongs. Columns 26 and 27 are used for storing information about homographs.[20] Column 26 indicates the type of the homograph, while column 27 indicates to how many form classes a homograph belongs. Thus the homograph что "that, what, which" belongs to four form classes, nominal, adverb, conjunction, and declined relative. Other morphological information contained in the grammar coding sheets pertains to gender, number, case, animation, tense, mood, voice, person, for inflected forms, while the noninflected forms contain information characteristic of them. Thus, for the adverb гораздо "considerably" there is an indication that it must be always followed by a comparative. In case of the prepositions, there is an indication as to whether they can be used postpositionally, like спустя "later," ради "for the sake of," вслед "after." Conjunctions are marked as to whether they are coordinating or subordinating, ambiguous or nonambiguous, paired or not paired.

Syntactic relationships like agreement and government are also coded wherever applicable. The agreement code indicates what the requirements are for the subject of a given predicative, or the possible cases, numbers, and genders which a nominal, a modifier, or a declined relative may have in all possible contexts. The government code is used to indicate which cases are governed by propositions. There are

[19] This classification scheme was first proposed by Paul L. Garvin, of the Bunker-Ramo group and later adopted, with modification, by the Wayne State group.

[20] A homograph in a given language is a word which, when projected into all possible contexts, may function in more than one part of speech category, e.g., English "hit" which may function as a noun, adjective, and verb.

indications of what specific cases are governed by prepositions, modifiers (аналогичный—"analogous"), and nouns (подарок брату—"a gift for brother"). For predicatives, infinitives, and gerunds, in addition to the indication of cases which they govern, there are also marked cases of what can be labeled complex government, which is said to exist when a predicative governs more than one case, one of which may be a prerequisite for the others: Его выбрали секретарём—"They elected him as a secretary," the instrumental is required when the accusative is present. Determination of the presence of agreement and government relationship is vitally important in grouping lexical items into sentence units, as will be shown later. Syntactic information about punctuation and formulas occurring in the text is also coded. For example, if formulas serve a predicative function in the sentence, they are appropriately coded in the predicative class.

Target language equivalents of source language dictionary entries are, of course, also entered. In the English IBM card of the Wayne group the first 24 columns contain the Russian form. Column 25 has an indication of the form class of the Russian form, column 20 has the number of translations, column 27 is blank, and columns 28-72 contain the English equivalents. All the above information is then put on magnetic tape and read into the core memory of the computer.

3.6.2 Wayne State *University Automatic Syntactic Analysis*

After the lexical units of the source language have been looked up in the automatic dictionary, properly identified, and supplied with all the stored grammatical information, syntactic analysis of the source language begins.

The Wayne group employs as the first phase of its syntactic analysis, procedures called blocking, the purpose of which is to identify certain sentence elements that are considered to be important, and to group them with certain adjacent elements that are considered to be linked to these elements in a complementary way. This blocking procedure is useful because it results in the creation of a sentence image, with fewer elements, which simplifies subsequent sentence analysis for the identification of main sentence units, subject, object, predicate.

Essentially, the blocking procedure changes the sequence of lexical units into a sequence of grammatically meaningful blocks (which may consist of just one word). Taken in isolation, without reference to the rest of the sentence, these blocks may still be ambiguous on this level, since the computer does not possess the speaker's intuitive ability to assign them their correct role in the sentence. The present Wayne procedure, called HYPERPARSE, permits the systematic establishment of a sentence profile (by proceeding from the blocking level to the phrase level, and from that to the clause level, each time utilizing the information derived from the previous level). This results in the identification of the kernel of the Russian sentence (subject, verb, object) and the adjunct elements related hierarchically to the kernel. On the phrase level, by means of appropriate routines, the blocks are first divided into four classes. The first class contains predicatives and all potential clause markers, such as conjunctions and punctuation marks. The second class contains potential candidates for subject and object, i.e., possible nominative and accusative constructions. Blocks identifiable as genitives constitute a third class, which serves as a transitional class between the potential kernel elements and adjuncts, while the fourth class contains all prepositional blocks, adverbs, all datives and instrumentals, and all other phrases which are identifiable as potential adjuncts.

After the blocks have been divided into the above four classes, the syntactic analysis on the clause level begins with the systematic investigation of the contents of the first class. This provides information about the structure of the entire sentence, and results in the selection of procedures for analysis of the membership of the other three classes, which are analyzed in turn. Agreement and government checks are used wherever appropriate in the course of analysis, the final result of which is the establishment of the constituents of the sentence in a hierarchical fashion. This analysis is designed specifically for the morphological and syntactic features of Russian, and, furthermore, its results provide an output from which the synthesis proceeds, and which involves, as pointed out above, a contrastive treatment of Russian and English.

3.7 Other MT Syntactic Analysis Techniques

Another type of syntactic analysis called predictive analysis was used by two machine translation groups in the United States: National Bureau of Standards and Harvard University. The notion of predictive analysis was introduced by Ida Rhodes of the National Bureau of Standards [*25*]. Her empirical system was later adopted, with modifications, by Anthony Oettinger and Murray Sherry at Harvard [*26*]. In predictive analysis, the syntactic analysis program scans a sentence from left to right, and a storage device called a "pushdown store" is utilized in the "prediction pool," i.e., the list of anticipated structures. The pushdown store is a linear array of storage elements in which information is entered and removed from one end only, so that the last element (prediction) entered is the first one picked up for testing. The technique of predictive analysis subsumes that, in scanning a

Russian sentence from left to right, it is possible both to make predictions about syntactic structures that occur to the right of any word, and to determine the syntactic role of a word itself on the basis of previously made predictions which the word might fulfill.

Mrs. Rhodes has developed a set of predictions describing Russian syntax, some of which are in the glossary and some of which are in the syntactic analysis routine. Each prediction has an urgency number. When a prediction is fulfilled, it is erased. If a prediction is unfulfilled and has a low urgency number, it is also erased. The unfulfilled predictions with high urgency numbers are kept until the end of the analysis, and are used as criteria for evaluating the quality of the analysis. The predictive analysis scheme of Mrs. Rhodes assumes a previous profiling of the sentence to determine clause and phrase boundaries.

Sherry's predictive analysis incorporated provisions for a single path only. His system was found to be inadequate for most sentences, especially for long ones, because of the impossibility of always choosing the correct alternative at each point, particularly in a sentence where syntactic ambiguities exist, and because of the difficulty in tracing back an error known to exist because of inconsistency at some later point in the analysis. In addition, the single path system did not indicate situations where more than one parsing of a sentence was possible. As a result, Warren Plath [*27*] of Harvard, basing his work on that of Susumo Kuno and Anthony Oettinger of Harvard (in connection with the syntactic analysis of English), developed a multiple-path system of predictive syntactic analysis, which investigates all possible parsings of a sentence, discards the paths which have little likelihood of occurring, and conveniently solves many of the earlier problems involving syntactic ambiguities and error tracing.

Still more research, both on the grammatical analysis of the source language and that of the target language, needs to be done before machine translation output begins to resemble that of a human translator. The contents of language lexicons and what has been referred to above, as contained in the head of the human translator, has to be stated in a form acceptable for computer processing. As an example of this type of endeavor, could be cited the research that the Wayne group is undertaking, for compiling the contents of two Russian lexicons for computer storage and retrieval.[21] To this store

[21] Such data were gathered with the support of the National Science Foundation through the activities of two research grants: *Comprehensive Electronic Data Processing of Two Russian Lexicons* and *Computer-Aided Linguistic Analysis of Russian Lexicon: Development of the Grammatical Profile of Lexical Entries.* The two Russian lexicons which served as the corpus for these investigations are: *Tolkovj slovar' russkogo jazyka,* Lexicon of the Russian Language (D.N.Ushakov, ed.) Moscow, 1935 and *Slovar' russkogo jazyka* Dictionary of the Russian Language Academy of Sciences, USSR.

of information should also be added knowledge about the language which still had not been codified, e.g., which Russian verbs must be always followed by a direct object (obligatory government). Some help for English synthesis is sure to come from attempts to construct generative transformational models of English grammar by disciples of Chomsky and Harris. Contributions in the area of semantics, which has barely begun to be investigated in machine translation, including the Bunker-Ramo [*28*] and RAND groups [*29*], and others [*31, 32*], are sure to come in the future. More investigations of language structure and the development of automatic syntactic recognition and sentence parsing routines are necessary for the implementation of effectively functioning MT systems.

4. Survey of MT Groups Since 1960

When Bar Hillel's contribution on the status of mechanical language translation was published in volume I of *Advances in Computers* in 1960, formally organized MT research activities had been established principally in three countries: the United States, the USSR, and Great Britain. In addition to his discussion of developments by the various MT groups in these countries, Bar-Hillel included some remarks on the status of MT research at Milan, Italy and the Hebrew University at Jerusalem. In the nearly 10 years that have followed Bar-Hillel's survey, reports of official research in MT have emanated from as many as seventy groups in at least fifteen countries: United States, USSR, Great .Britain, France, Germany, Japan, Italy, Belgium, Canada, Mexico, Czechoslovakia, Yugoslavia, Rumania, Hungary, and Poland. If efforts of directly related research in computational linguistics or hardware development were included, these figures would be greatly increased.

Moscow, 1957. The main objectives of these investigations were (1) to analyze and process for computer retrieval the pertinent linguistic information contained both explicitly and implicitly in the above lexicons; (2) to compile information contained in each of the dictionaries in the following broad areas: lexical, morphological, and that of function and usage, and then to merge this information; (3) to treat various methodological and theoretical approaches or lexicological research. This information, it is hoped, will provide a useful and rapidly accessible data base for further processing and utilization by grammarians as well as computational linguists. Based on this research, a series of publications was planned, the first of which - *Distribution of Ten Canonical Entry Classes in Two Russian Lexicons* appeared in January 1967. The second publication in this series by Alexander Viteie, *Russian Substantives: Distribution of Gender,* appeared in the summer of 1969, and deals with the distribution of gender among Russian substantives in the above lexicons.

The present survey will first discuss briefly the status of each of the principal groups covered in Bar-Hillel's survey. Second, several groups which came into being after the publication of Bar-Hillel's article are surveyed. Third, in contradistinction to all the above groups which are essentially research-oriented, the current activities of five production-oriented MT projects are described in some detail.

4.1 MT Groups Covered in Bar-Hillel's Survey of 1960

4.1.1 University of Washington

The Seattle group investigated problems of MT from 1949-1962. Primary emphasis was placed originally on German--English MT, and subsequently on Russian-English MT, but a separate program was also undertaken for Chinese-English MT. At the time of termination, advances in MT research at Seattle included considerable development; of dictionary procedures its well as morphological and syntactic analysis routines. A particular effort was made in investigating the problem of automatic resolution of grammatical ambiguity. At the conclusion of this research, machine translation was being simulated as a means for pinpointing system procedures which required improvement, e.g., codification of lexicographical or dictionary data.

In addition to the development of an MT system, the Seattle group conducted collateral research in translation evaluation in investigations of (1) the statistical aspects in translation of scientific language, and (2) machine translation quality and its evaluation.

The principal goals of the Seattle group according to its view of the translation process were (1) selection of the correct target equivalent of the source word, and (2) organization and coordination of these words into a grammatical and semantic form which would be meaningful to the reader in the target language.

To implement the group's evaluation program, three general approaches were investigated:

(1) A study and evaluation of occurrence probability between certain Russian kernel structures and their English equivalents

(2) A study of methods evaluating continuity of meaning among equivalents selected by the translation process

(3) The development and evaluation of a "test" of translation reality.

The Seattle approach of kernel analysis entailed the mechanical location of basic elements in each sentence of a Russian text. These elements are classified according to types based on their grammatical features and word order. English kernels, composed of words in the English translation of the text corresponding to the words of the

Russian kernel, are also mechanically located and classified according to type [*33-35*].

4.1.2 Massachusetts Institute of Technology

The, M.I.T. group was engaged in formal MT research for 12 years, 1953-1965. Research efforts were primarily devoted to German- and Russian-English MT; other languages, such as Arabic and French, were also investigated. Since it was felt that for its objectives both source and target languages required complete analysis, the M.I.T group performed extensive analysis of English as well.

The basic approach to grammatical analysis of this group was based on the phrase structure model. To meet programming needs a new program language was developed by the principal investigator, Victor Yngve. This language, known as COMIT, gained wide acceptance and usage by many MT workers. Various investigations ranged from programming techniques to theoretical linguistics, from morphology and syntax to semantics. An over-all aim of fully automatic quality translation led the group throughout its research period to investigate many problems irrespective of degrees of complexity. A very impressive list of publications issuing from its research appears in the *Final Report* [*36*].

4.1.3 Georgetown University

In his 1960 survey, Bar-Hillel said of Georgetown that no other MT group has been working on such a broad front. In the report on Georgetown's activities from 1952-1963 prepared by R. Ross Macdonald, the full scope of MT research, together with a complete list of publications for that period, is presented in detail.

The Georgetown group committed itself to result-oriented experimental research in primarily Russian-English MT of scientific and technological literature. Since the basic problem of MT was considered to be linguistic in nature, the Georgetown researchers were mainly guided by current linguistic orientation.

The group's main objectives, as stated in the above report, were the following:

(1) The progressive improvement of experimental runs by means of a feedback procedure on the basis of the discernment of lacunae and inadequacies

(2) The processing of scientific, and technological language first of all, since it presents fewer problems than general language for automatic translation

(3) Transfer from Russian to English.

Following the publication of the above report, individual papers by various group members have appeared under the series title *Occasional Papers on Machine Translation,* see [*37*]. These papers are numbered consecutively and are available at Georgetown University. They cover a wide range of topics on materials which were generally developed around the Georgetown technique of structural transfer.

Since 1965 MT research has been continued at Georgetown under the direction of R. Ross Macdonald. The objective of this research has been to achieve an operational MT system which will transfer Russian to English without editing, and which will be usable in connection with any effective general purpose computer.

4.1.4 RAND Corporation

The RAND group began its studies in machine translation around 1958 with primary emphasis being placed on Russian-English MT until approximately 1963, when the scope of research was broadened to include problems in theoretical and structural linguistics. The approach to MT taken by this group was based on dependency theory as elaborated by David G. Hays. This method, which Hays called "sentence structure determination," sought to establish dependency relationships among text occurrences in the sentence. Dependency theory, according to Hays, is actually a characterization theory not necessarily associable with any empirical method or principle. It is a theory of grammars with abstract mechanisms for characterizing sets of utterances, and for assigning to them certain structural descriptions called D-trees.

The activities of the RAND group have centered generally around computational linguistic research. The group has produced a great quantity of documentation which has been released periodically over the past 10 years in the form of monographs. Its contribution to machine translation research was indeed of considerable scope and influence [*38-41*].

Two additional efforts reported in 1966 were studies on predication in contemporary standard Russian, and elaboration of the RAND glossary of Russian physics.

4.1.5 Bunker-Ramo Corporation

This is fundamentally the same group as comprised the Ramo-Wooldridge group (as reported in Bar-Hillel's survey) and the Thompson-Ramo-Woolridge group. To clear up any confusion, the successive changes in the name of the corporation did not involve either geographic displacement of the group or significant alteration of its research objectives and personnel. The large number of progress reports and other publications since at least 1958 have of course appeared under each of the above names. MT research has centered chiefly around Russian-English MT, but there has been extensive work with other languages as well.

The basic approach of this group centers around the fulcrum theory as enunciated and elaborated by Paul Garvin. Most recent fulcrum techniques have been concentrated significantly on the resolution of semantic problems.

The major research effort at Bunker-Ramo led to the production of an experimental but operational machine translation program which included input, dictionary look-up, syntactic analysis, and printouts of translation of scientific documents as well as selected text from the Soviet newspaper *Pravda.*

Like other MT groups, such as Wayne State University, Texas, and RAND, the Bunker-Ramo group, through its experience, reached the conclusion that high quality machine translation is still not feasible, such output being impeded by many unsolved problems. These problems are linguistic as well as computational. The hope was that the quality of MT output will increase proportionally as these problems are articulated and resolved.

MT research at Bunker-Ramo has involved four major areas:

(1) Linguistic analysis leading to more effective syntactic analysis of Russian

(2) Linguistic analysis directed at improving the quality of the English output

(3) Incorporation of the results of the linguistic research

(4) A continuing programming effort directed at improving the simplicity, dependability, and the speed of the computational aspects of the system.

Although, as mentioned above, the Bunker-Ramo group has been involved recently in deep semantic studies and related problems of multiple meaning, considerable emphasis was also placed on syntactic analysis, specifically Paul Garvin's principle of a heuristic approach to syntax. In this approach, decisions are based on immediate context as tentative, subject to revision on the basis of information from a larger context. In addition to various progress reports, see [*12-16*].

4.1.6 Harvard University

As pointed out in Bar-Hillel's survey, the Harvard group's primary MT research effort in the late fifties was devoted to the development and implementation of an automatic dictionary, an MT problem to which Harvard, perhaps more than any other research group, devoted considerable experimentation. During that time, the source language was Russian.

By 1960, this group shifted its emphasis from word-by-word translation to sentence analysis. Concomitantly, grammatical analysis of Russian structure was implemented on a broad plane, and an approach to the automatic syntactic analysis of the sentence was formulated. This approach was the predictive method originally formulated by Ida Rhodes, and modified and adapted by the Harvard group. It was tested on English sentences during initial stages, and gradually was extended to Russian as well. The general strategy employed in syntactic analysis came to be known as multiple-path syntactic analysis.

Since 1964, the Harvard group has expanded its investigations to include mathematical models of language in addition to research on automatic syntactic analysis of Russian and English [*47-51*].

4.1.7 University of Michigan

Research in machine translation by this group was discontinued in 1962. The primary research objective at that time was the development of a theory of translation to form the basis for constructing a "learning" or "self-modifying" program to translate natural languages. See Koutsoudas [*52*] and Smoke and Dubinsky [*53*].

4.1.8 University of Pennsylvania

MT *per se* was not the objective of this research group, whose principal investigator, Zelig Harris, was the prime author of the transformation model of language analysis.

The principal aim of this group was automatic syntactic analysis. In the course of its development, the group contemplated computerized syntactic recognition. In this connection, there was some tangential bearing on MT, insofar as this group, as well as various MT researchers, concerned itself with the automatic generation of English output.[22]

4.1.9 National Bureau of Standards

Bar-Hillel briefly described a method of syntactic analysis of Russian devised by Ida Rhodes and applied by the NBS group in their MT experimentation. This method later became known as predictive

[22] See various papers printed under series heading University of Pennsylvania, Transformations and Discourse Analysis Projects, particularly Sager, Naomi, "Procedures for left-to-right recognition of sentence structure,'" No. 27, 1960; Harris, Zelig, "Immediate constituent formulation of English syntax" No. 45, April 1963; Holt, Anatol W., "Mathematical and applied investigation of tree structures of computer syntactic analysis," No. 49, June 1963.

analysis, which, as mentioned in the discussion of the Harvard Group, was later adapted and modified by Harvard.

MT research by this group continued until 1964. At that time, work was proceeding on the last of three sections in the long-range program for practical Russian-English MT. The first two sections dealt essentially with (1) the mechanical retrieval of source word grammatical characteristics, and (2) the parsing of Russian sentences using the predictive method. The last section involved a special process of encoding of words called "profiling" [*54, 55*].

4.1.10 Wayne State University

Research in computer-aided Russian-English machine translation at Wayne State University has led to the development of an experimental system which presently comprises three basic operations—dictionary look-up, syntactic blocking, and HYPERPARSE. At the current stage of development, these three operations are coordinated through the interaction of both human and machine procedures. (See Fig. 1.)

Input to the system generally consists of a technical Russian language text which has been keypunched and read onto magnetic tape. One sentence at a. time is analyzed. Each sentence item is looked up in a computer-stored dictionary (a compilation of words as they occur in a text comparable to the input text) and its encoded grammatical characteristics and English equivalent(s) are retrieved. The resultant output tape of looked-up text is used as input to the next operational phase—syntactic blocking or analysis. This procedure endeavors to recognize and to record the functional (grammatical) role played by each sentence constituent.

In syntactic analysis, sentence items (sometimes only one word) which serve the same function in the sentence, e.g., subject, predicate, object, are put into blocks or groups. Each block has a kernel word (noun, verb, preposition) on which other sentence items may be said to depend. For example, blocks would be formed by a noun and its modifiers, a verb and an adverb, a preposition and its object noun, etc. An output tape of blocked sentence is then created and serves as input to the third operational stage, i.e., where the blocked sentence is automatically parsed. This procedure entails determining the functional role (e.g., subject, predicate, object) played by each of the blocks.

To resolve the problem of grammatically ambiguous sentences, the computer program HYPERPARSE (see Section 3.6.2) reduces grammatically ambiguous sentences to the fewest possible interpretations. This is done by means of a mechanically generated matrix which seeks to discover distinctions among the apparently ambiguous blocks. Those noun blocks which do not qualify as either subject or object candidates are grouped as so-called "adjuncts."

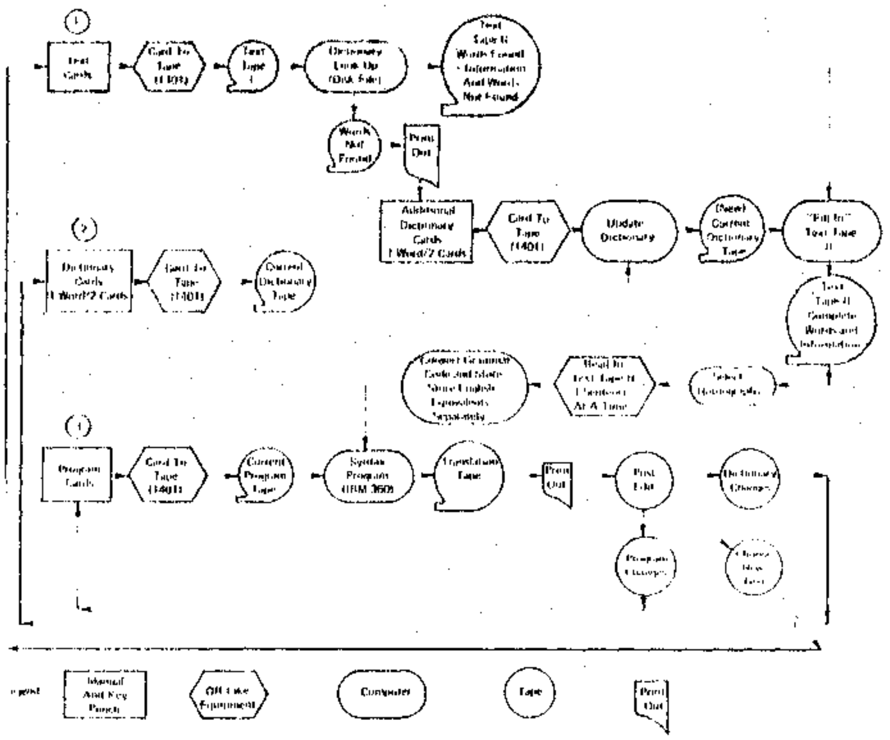

FIG. 1

Experimentation in machine translation at Wayne State University[23] is continuing with two primary objectives: first, to automate additional analytic procedures; and second, to refine previous routines according to insights gathered from each successive experiment. At the same time, translation rules are being formulated which, when written into the system, will improve the quality of the output. For a complete list of all MT publications produced between 1958 and 1968, see [*58*].

For one year's research, Shogo Ushiyama, from Japan has joined the WSU group. The objective of his research is the design of a system for automatic translation of Japanese- English telegrams.

[23] For dictionary problems and grammar coding see [*56*]; for a discussion of problems in automatic syntactic recognition, see Steiger [*57*].

4.1.11 University of California at Berkeley

The particular language emphasis in MT research by this group has been on Russian- and Chinese-English. Extensive research in theoretical studies of linguistic structure has been conducted. It has been the belief of this group that these studies would yield formalisms which might serve as a basis for writing MT algorithms.

The grammatical approach taken by this group was significantly influenced by the stratificational model developed by Sydney Lamb while he was at Berkeley. In his syntactic research, he was primarily interested in developing a system for tactic analysis which presupposes a segmentation of text units in terms of his concept of language as a hierarchy of strata.

A wide range of linguistic problem solving by computer has characterized research during and following the period of primary concentration on. the problem of MT. Developments growing out of this research included such studies as production of concordances by computer. For a discussion of Lamb's approach to automatic syntactic analysis, and linguistic models see [*59-61*]. Efforts to develop a Chinese-English MT system have been continuing under the, direction of Ching-Yi Dougherty.

4.1.12 University of Texas

This group documented its activities from 1959 - 1961 in a series of quarterly progress reports entitled "Machine Language Translation, Study." Since 1962 its activities have been incorporated into the Linguistic Research Center at the University of Texas, under which title group publications have appeared up to the present time.

This research group has always had a slightly different orientation than most other MT research groups in the United States owing to the fact that attention has been focused not only on problems of machine translation, but simultaneously on the general area, of linguistic analysis. Efforts have centered around three interrelated tasks directed at refinement of the group's model and approach. These three tasks are (1) descriptive linguistics, (2) programming, and (3) mathematical research.

As reported recently in a pamphlet from the Linguistic Research Center, the approach is concentrated on four points:

(1) The translation process is viewed as independent of specific source and target languages.

(2) The structures of source and target languages are supplied to the computer in the form of grammar codes.

(3) Linguistic analysis produces phrase structure grammars arranged

in a hierarchy (i.e., independent but related grammars are produced for each type of description, e.g., for syntactic and semantic).

(4) Linguistic analysis of source and target languages are independent of each other.

The Linguistic Research Center approach, although geared at the outset to German-English translation, has aimed at developing a translation technique that could be readily adapted to other languages. The group has, in fact, been developing syntactic descriptions of English, Russian, and German, and limited efforts have been made for similar descriptions of Japanese and Chinese. Pilot descriptions of other languages have been made as well to test the generality of the algorithms developed for machine translation.

The group has sought to accumulate comprehensive data including lexical, syntactic, and semantic descriptions of many languages. These data are then retrieved in compact formats required for automatic analysis and synthesis. Around 1965 the Russian Master Dictionary, developed through Air Force support, was incorporated into the research program. The dictionary has permitted the testing and expansion of a Russian-English translation system based on a syntactic translation model that has been developed [*62-64*].

4.1.13 Birkbeck College, London

As was clearly pointed out in Bar-Hillel's survey, Andrew Booth, the director of this group, was one of the outstanding MT pioneers. It was also indicated that this group placed great hope for successful MT on the development of an interlingua, i.e., an artificial intermediate language: moreover, the group was guided in its experimentation by methods based on mathematical and logical considerations.

Research into MT techniques continued into the mid-sixties. However, there was more progress made in a related area of investigation, viz., the mechanization of certain aspects of literary studies, such as the preparation of word-indexes and concordances. For a sample of this group's work see [*65*].

4.1.14 Cambridge Language Research Unit, England

It was shortly after the publication of Bar-Hillel's survey, in which he briefly described the newly conceived CLRU approach to grammatical analysis, that there was a final formulation of its lattice theory on which the CLRU syntactic analysis was subsequently based. In addition to syntactic analysis, the CLRU also focused attention on development of its thesaurus approach as well as on an investigation of semantics in general.

Its research progress was reported regularly until approximately 1965. Specifically, during the time between the appearance of Bar-Hillel's survey and 1965, its work in syntactic analysis entailed the development of an automatic parsing program based on the aforementioned Parker-Rhodes lattice model of language. A formal bracketing (i.e., syntactic blocking) algorithm was developed. For a formal presentation of the lattice theory approach see [*66*].

In the area of semantics, CLRU research activities included the development of a general interlingual machine translation procedure, and an investigation of the semantics of language based on the "semantic square" model of human communication. A detailed view of the CLRU semantic approach and related discussions may be found in [*67*]. Margaret Masterman of this research group has served as a consultant to the National Research Council of Canada on problems relating to development of MT in that country. Most recently reported research was focused on English to French machine-aided translation. The group is attempting to develop a system for producing low-level but message-preserving output, based on segmentation of text into natural phrasings (i.e., units of text containing two stress points, or a stress point and a pause), by computer program.

4.1.15 MT in the USSR

In his survey, Bar-Hillel included a discussion of the MT research activities of approximately ten groups which were conducting research at that time in the USSR. He indicated that whereas some groups were theoretically oriented, others were more empirically minded; some advocated FAHQT, while others operated with more modest aims. Not unlike MT in the United States, serious research in the USSR shifted its emphasis and came gradually to recognize the more realistic aim of machine-aided translation. With a clearer understanding of the problems of MT, serious researchers were able to emphasize the need for descriptive studies of languages "from a structuralist point of view" as Bar-Hillel pointed out, and "the only one which makes sense for MT."

Rather than undertake a similar survey here, since the pertinent literature is not so readily available, a comprehensive summary of the state of the art in the USSR, is recommended for the interested reader. It has been translated from the Russian and contains a bibliography of 165 significant references [*68*].

A portion of the closing statement perhaps best sums up the current view of MT in the USSR, and it bears quoting:

> The first results of work on machine translation created an illusion in the scientific community that the solution to the problem was near at hand. This illusion was due not only to the seemingly convincing demonstrations of machine translation capabilities. . . . Nor did these demonstrations have any significant impact in the scientific community. Much more serious is another illusion which was supported by many investigators. This is the illusion that the problem at hand was of an "engineering" nature, that machine translation had already been resolved in principle, and that to be implemented in practice, only considerable organizational efforts were required ([*68*], p. 43).

The authors went on to add:

> An opinion was maintained that large-scale machine experiments with translation using known principles would be capable of making the available theoretical algorithms operational. It turned out, however, that the problem was in fact of a different class, for it concerned the investigation of operating principles of a biological system and the elucidation of how such a complex system as human language is capable of functioning at all. This is a serious problem indeed. Unless it is settled, no successful development of information systems needed for human society seems possible ([*68*], p.44)

4.1.16 University of Milan, Italy

Although the Milan approach of correlational grammar mentioned above was considered unpromising by Bar-Hillel, it endured at least until 1966 at Milan, and has provided a basis for further research at the University of Georgia on the same theoretical model.

Although basically a theoretical approach to MT was favored by the Milan group, many practical activities were undertaken. The group worked primarily on Russian-English translation, but German and Italian output were added to the procedure. An experiment of English-Latin translation of random sentences made with a limited vocabulary was also perfected. Further efforts included the development of a general grammar suitable for the mechanization of every kind of linguistic activity: translation, summarization, and description. For a general view of the Milan effort see Ceccato and Zonta [*69*] and Ceccato [*70*]; see also Ceccato [*71*].

4.2 MT Groups Established After Bar-Hillel's Survey of 1960

These groups, for the most part, are located in countries other than the United States, the USSR, Great Britain, and Italy.

Inasmuch as there was relatively little literature on the activities of these groups available at the time of preparation of this survey, the following discussions are quite brief. The omission of some existing groups from this survey is in no way intended to give the impression that the activities of these groups do not merit mention. It simply means that no pertinent material relating to their work was accessible,

4.2.1 Centre d'Etudes pour la Traduction Automatique, University of Grenoble, France

This research group has been conducting experimentation in Russian French machine translation for the past few years. Having first compiled a machine dictionary, the group then proceeded to write a, formal recognition grammar for Russian syntactic structures. The transformation model of grammar has been adopted in attempt to develop a metalanguage common to the two main procedural problems: (1) semantic evaluation of syntactic structure in the source language, and (2) generation of French syntactic structures. The group has recently been testing transformation rules written in conjunction with the above grammar for analysis of tree structure.[24]

The group has also been experimenting with German-French and Japanese-French machine translation. Further information is available [*72-75*].

4.2.2 University of Montreal, Canada

The Department of Applied Linguistics of the University of Montreal has conducted a mechanical translation research project since 1966, sponsored by the Canadian National Research Council. Research efforts have been focused on English--French machine translation. In organizing its program this group maintained close contacts with automatic translation groups at Harvard University and the University of Grenoble, France.

The primary aim, following the compilation of a dictionary in which morphological information is stored, has been to compose a grammar of English in order to be able to perform automatic recognition of syntactic structures. Existing algorithms have been tested to study the comparative efficiency of the respective syntactic analysis systems. The group appears to have expressed a preference for tree structure analysis and is continuing to experiment with grammatical models (e.g., finite state grammar). In addition to the quarterly progress reports of this group, see the following [*76-77*].

[24] A recent report indicates that a stratificational approach has also been incorporated into this group's efforts to design an MT system.

4.2.3 Groupe de Linguistique Automatique, Université Libre de Bruxelles, Belgium

This group, headed by Lydia Hirschberg, has worked in cooperation with EURATOM. The analytic procedure has many points of similarity with Zelig Harris' string analysis. The group has also patterned its analytic procedures to some extent on Hays' dependency grammar and its related tree structure. Having developed a morphologically coded dictionary and syntactic analysis procedures, the group, according to its most recent publications, has been investigating semantic problems, e.g., the use of semantic information in the selection of lexical units in microglossaries.

Since at least 1964, this group has also been developing automatic dictionaries for human translators. The group has been working in cooperation with its primary user, the European Coal and Steel Authority. Dictionaries of technical terms and idiomatic structures have been compiled with a view to translation among the four languages of the European Community: French, German, Dutch, and Italian. Dictionary look-up programs accept input sentences, and produce a listing of the vocabulary requested by a translator. There are also programmed provisions for morphological analysis of the source language. Among the various publications of this group, see in particular [*78*].

4.2.4 Karlova University, Prague, Czechoslovakia

The objective of this group has been the investigation of the Czech and English languages for machine translation. The group's analysis of English has been largely based on a modified version of the predictive method of analysis. In a related project, this group has been developing algorithms for Czech in connection with the generative description of the language. Work has also been done on the synthesis of the Czech language, grammar coding programs have been designed, and an intermediate language, recently developed, is being improved [*79*].

4.2.5 Computing Centre of the Hungarian Academy of Sciences

The research objective of this group is automatic syntactic recognition of Russian sentences. A procedure called pseudomorphemic syntactic analysis has been developed which is designed to control and minimize the number of grammatical rules to be written. This scheme includes morphological analysis that provides input for syntactic analysis in the form of initial sequences of categories. The number of rules required for these analyses is reduced since separate categories are set up not only for word forms, but for their grammatical properties as well [*80*].

4.2.6 Deutsche Academie der Wissenschaften zu West Berlin, Germany

This group has been working on English-German and German-English MT. Its objectives include research on mathematical foundations, development of methodology, and establishment of algorithms as scientific preparation for practically applicable English and Russian machine translation of natural science and technical texts by means of computer techniques. In addition to other group reports, see [*81*].

4.2.7 Japan

Research in machine translation has also been based at several projects in Japan, including the following four groups: (1) Electrotechnical Laboratory, Tokyo, Japan; (2) First Research Center, Defense Agency of Japan, Tokyo, Japan; (3) Kyoto University, Kyoto, Japan; (4) Kyushu University, Fukuoka, Japan.

Group 1 has concerned itself with English-Japanese and Japanese-English translation of scientific papers. A generalized and syntax-oriented translation system is its objective. Research has included programming considerations as well as formal descriptive linguistics. For additional information see [*82*].

Group 2 has also concentrated on problems of syntax. Its approach seeks to help linguists construct a less redundant set of rules, and also to give comprehensive account of the algorithms for syntactic analysis, transfer, and synthesis. See [*83*].

Group 3 has a translation procedure from English to Japanese, the principal feature of which is the use of four grammar tables. Analysis is done from the end of a sentence to the beginning, applying these four tables recursively. See [*84*].

Group 4 has been placing current emphasis on the development of a practical procedure for reciprocal English-Japanese translation. Work falls into three areas: (a) the accumulation, analysis, and organization of linguistic information for constructing dictionaries and formalizing Japanese grammar; (b) computer programming; and (c) the introduction of a learning process to the computer program of translation for the automatic analysis of syntax. See [*85*].

4.2.8 Other MT Groups

Research in MT has been conducted at the following centers in some cases since the early sixties. The lack of up-to-date information on the activities of these groups has made it difficult to summarize the results of their research. The names of some of these groups are being listed here so that the reader may at least be cognizant of their existence.

Universidad Nacional Autónoma de México, Mexico
Université de Nancy, France
University of Debrecen, Hungary
University of Saskatchewan, Canada
Research Institute for Mathematical Machines, Prague, Czechoslovakia
University of Warsaw, Poland
Institut Za Eksperimentalnu Fonetiku, Yugoslavia
Académie de la République Populaire Roumaine, Rumania

4.3 Production-Oriented MT Groups

In contract to the foregoing discussions of the activities of research-oriented MT research groups, this part of the survey will describe the activities of organizations that have committed themselves to the development and operation of production-oriented machine translation systems.

4.3.1 European Atomic Energy Community (EURATOM)

This organization is located in Ispra (Varese) Italy. The principal director of the Russian-English machine translation program is S. Perschke. This center, which has adopted the machine translation system developed at Georgetown University, has sought primarily to carry out practical machine translation of Russian scientific literature.

The translation system is reported to be operative and to be producing quite acceptable scientific translations. Whenever errors are discovered in the output, the dictionary and linguistic operations are updated.

Improvements in the system were contemplated within the past two years, including new procedures for syntactic and semantic analysis [*86*].

4.3.2 Oak Ridge National Laboratory, Oak Ridge, Tennessee

The machine translation system at Oak Ridge has been essentially adapted from the system developed through experimentation by the Georgetown University machine translation project. The Oak Ridge program pursues two objectives: (1) to provide rapid translation service to scientists and engineers; and (2) to continue MT research in order to improve the existing system. The chief advantage of the system is the ability to produce translations rapidly. Conventional translation by an outside agency requires 4 to 6 weeks, whereas the MT output is obtainable in 1 week, and, if there is urgent need, even overnight service is possible. The cost of translation has turned out to be competitive with human translation.

The group has not yet conducted a rigorous evaluation of the comparative quality of MT output as opposed to human output. However, expressions of opinion have been solicited from many users, and with little exception opinions have been favorable.

In connection with the group's second objective, there has been a concentration of effort on introducing new words into the machine dictionary on the basis of actual texts. Refinement of translation equivalents is also planned, and it is hoped that microthesauri in specialized fields will be developed.[25]

4.3.3 Foreign Technology Division, U.S. Air Force, Dayton, Ohio

This MT system has been operational for several years, and is based primarily on a dictionary developed by IBM and on a grammar worked out by a research group at the University of Texas, Austin, Science text is processed from Russian to English. The output is examined by human post-editors before it is transmitted to user agencies.

An evaluation of machine aided translation activities at the Foreign Technology Division was made in 1965 in terms of quality, cost, and timeliness [*87*]. The quality of the output of the machine-aided translation system, for the group's purposes, was considered comparable to standard human translations. The main cost components in the process, i.e., post-editing and recomposition, accounts for 70% of the cost, but over-all comparative costs between human and machine-aided translation were competitive. In terms of time spent, it was found that machine output often required less than half the amount of time for human translation.

It was concluded that the "MT system in both is technically and economically feasible," and that it should be continued and operated as a production facility devoted to providing high-quality service to users of translations.

4.3.4 National Physical Laboratory, Teddington, England

Research in MT which has been in progress for several years at this facility was recently concluded (see McDaniel *et al.* [*88*].The primary aim of this group had been to demonstrate the practicability of translation

[25] Information about activities at Oak Ridge was obtained through personal correspondence between the author and Francois Kertesz of the Oak Ridge group.

by computer of Russian scientific and technical texts in the fields of electrical engineering, mathematics and physics into English. Experimentation and goals have been directed toward the simulation of a translation service, and in this respect the most recent efforts may be construed as operational machine translation techniques. The ultimate experiment involved a procedure for evaluation of the usefulness of machine output which was not subjected to human post-editorial treatment. This evaluation experiment, in which 34 evaluators commented on 19 distinct papers, has a range scale from 8 down to 0, viz. from "Fully adequate" to "mostly very good" to "fair" to "poor" to "useless." A majority verdict which was produced may be paraphrased as "mostly good enough, with a few obscurities."

A broader evaluation exercise has been contemplated which would be conducted over a wider range of potential readers, and which, it is hoped, might lead to a strengthening of the above verdict, thus making it possible to decide on the viability of a production machine translation service based on the NPL system.

It is felt that many components of the NPL system have been developed so that each may be of considerable independent importance. For example, the automatic Russian-English dictionary, covering all forms of around 17,000 Russian words, and available on punched cards, is perhaps capable of wider application. Included among other components of similar value are the scheme of comprehensive morphological representation, methods of Russian syntactic analysis, and a computer model of linguistic structure.

4.3.5 Central Research Institute for Patent Information, Moscow, USSR

Since 1964 this agency has been involved in developing and operating a system of "automatic translation" of patent literature from English into Russian. All programming has been done, on the Soviet URAL-4 computer; source text has been taken from the *Official Gazette,* the weekly publication of the United States Patent Bureau.

This system is quite sophisticated and therefore it may be profitable to examine it to see how at least one operational MT system is conceptualised and structured.

An algorithm based on segment analysis provides for the delineation of operational units of text (syntagmas) such as noun groups and verbal combinations. The translation is carried out with the aid of a compiled dictionary of specified patterns of syntactic constructions. In contrast to patent formulas, trade marks are not compiled in the text, but are examined separately.

The program for this system is comprised of approximately 20,000 instructions and consists of sixteen subroutines, enumerated below:

I. Text preparation
1. program to arrange words
2. program to search for words in the dictionary
3. program to analyze unknown words (i.e., not found in the dictionary)
4. program to process idioms
5. program for homograph selection
6. program to segment text
7. program to segment text into phrases

II. Syntactic analysis of segments
1. program for locating pronominal antecedents
2. program for working out case information (morphological)
3. program for analysis of predicative units of text
4. program for analysis of noun word combinations

III. Synthesis of Russian text
1. program for synthesis of Russian text and alphabetic printout of translation

IV. Auxiliary programs
1. master program
2. program for writing information on magnetic tape
3. program to transfer informal ion from tape to drum storage
4. program for printout of intermediate program results

For details see Kravec *et al.* [*89*].

5. ALPAC Report

5.1 Industry's Exaggerated Claims for MT Success

As evidenced from the above discussion, by the mid sixties most MT researchers were beginning to move along the path of learning how language operates and of attempting to incorporate the results of their research into the design of automatic procedures for the translating processes. There were in the United States, however, a few groups (notably Bunker-Ramo, Georgetown University, and IBM) whose primary aim in addition to the above objectives, was the development of a functioning MT system. Industry also jumped on the bandwagon, aiming primarily at developing a marketable MT system which would serve clients in need of translation. Its exaggerated claims for success raised hopes, later proved to be premature, and presented a false picture both to the general public and to government agencies, which by 1965 had given support to what was presented as MT to an amount close to $20,000,000 [*90*]. James P. Titus [*91*] gives the following very graphic description of industry's deep, but, as it turned out, rather misdirected interest in MT:

In industry an entirely different picture presents itself. In the early days—ten years ago—optimism for machine translation ran at a high pitch. Advertising, which eventually damaged the cause, glowed with promises of quick, clean translations. Marketing plans were laid for such adventures as automatic translating service centers, and one was opened by Itek Corporation in New York City. But it closed in a few mouths. Gradually, industry's enthusiasm for machine translation dwindled until it was either abandoned or submerged in other linguistic research.

IBM was one of the most, enthusiastic supporters of machine translation ten years ago, and it had a considerable effort under way. With funds of its own and funds of the government, it built four translation machines based on photo-storage and special-purpose, lexical processors: the Mark I, Mark II, the Research Language Processors — which was used at the 1964 World's Fair, and Alps, the Automatic Language Processing System. Two of these machines, Alps and Mark II, are still operating in the government.

5.2 Establishment of ALPAC

These exaggerated advertising claims, as well as the previously unheard of large expenditures of government funds in such a comparatively esoteric area as language, led in October, 1963 to the request by Leland Haworth, Director of the National Science Foundation to Frederick Seitz, President of the National Academy of Sciences:

> . . . to advise the Department of Defense, the Central Intelligence Agency and the National Science Foundation on research and development in the general field of mechanical translation of foreign languages [*90*].

Replying to this request, Seitz, in April, 1964 appointed the Automatic Language Processing Committee (ALPAC).[26] The committee carried out its investigation by examining the following three areas: (1) the need for translation in government agencies and the scientific community: (2) the satisfying of these needs by the then existing human translation facilities, government and private; and (3) the advantages, shortcomings, and perspectives for machine translation, including comparison between human translation product, and MT output — in terms of both quality and cost. In the course of its investigation the committee conducted a number of studies trying to evaluate the quality of translations, cost estimates, types of errors discovered in translation output, and other aspects of translation in general. It also interviewed seventeen witnesses, including translators, linguists, industry representatives, translation output users, and other invited people.[27]

[26] For the membership of the committee consult the appropriate page of the ALPAC report.

[27] For a list of persons who appeared before the committee see the ALPAC report, p.24.

5.3 ALPAC's Recommendations

The ALPAC report, entitled *Language and Machines* [*90*] was released to the public in November, 1966 and recommended expenditures in two distinct areas: (1) computational linguistics, and (2) improvement of translation. It also suggested by inference that pursuit of FAHQT is not a realistic goal in the immediate future, as reported in *the Finite String:*

> The committee sees, however, little justification at present for massive support of machine translation *per se,* finding it—overall—slower, less accurate and more costly than that provided by the human translator [*92*].

The committee also finds that

> . . , without recourse to human translation or editing. . . . there has been no machine translation of general scientific text, and none is in immediate prospect ([*90*], p. 19).

5.4 Reactions to ALPAC Report

5.4.1 German Reaction to ALPAC Report

Probably one of the most incisive points of the comments directed at the ALPAC report is, in the view of its critics, the contention that the committee's findings are based on a comparison of outdated machine translation output of 1964 vintage, ignoring more recent, improved MT output, which was then generally available, with products of human translators, some of whom were highly skilled professionals. Also ignored was the fact, in view of some, that if it is assumed that MT and human translation serve two different objectives the former of transmitting essential information, the latter of providing a full translation of a given text—then a completely different picture emerges [*93*]. The same critic, Friedrich Krollman, Director of the Translation Service of the Federal Armed Forces of West Germany at Mannheim, calls attention to the fact that the overwhelming part of the ALPAC report deals with the general problems of translation, such as quality, cost, service; the number, background, and availability of human translators; the rise of English as a language of science; the prospects of American scientists learning Russian—factors which relate, at best, only peripherally to MT. He finds it most astonishing that of the total 124 pages of the ALPAC report only six deal with the state of MT at that time. Of these six, only two-and-a-half pages discuss MT *per se*, with the remainder devoted to the reproduction of samples of MT output. Krollman thinks that more space in the report should have been devoted to the discussion of MT, its scope, objectives, and problems [*94*].

5.4.2 Soviet Reaction to ALPAC Report

Krollman [95] also questions whether ". . . the sometimes rather categorical conclusions of the ALPAC committee are valid for all future times and for all kinds of translation."[28] A similar view is also echoed in a review of the ALPAC report which appeared in the authoritative Soviet journal dealing with information processing generally, *Nauchno-Tekhnicheskaja Informatsija* (Scientific-Technical Information) which declares that while the findings of the committee may be applicable to the conditions in the United States, they certainly are not relevant to the Soviet Union and, at any rate, the state of the MT art in the USSR should also be investigated by an appropriate authoritative committee. The Soviet journal also finds that the ALPAC group has underestimated the progress of MT research and makes the following observation [*96*]:

It just seems to us that in the evaluation of the importance of this [MT] research the Committee displayed—perhaps in a most vexatious manner—a certain narrowness and pragmatic single mindedness, which is characteristic of the report. Those ideas which have originated and are originating in connection with MT are a contribution not only to the development of a MT system (a problem which is probably not acute in the United States) but also advance the resolution of one of the most important problems of the 20th century—the problems of symbiosis of man and machines.[29]

5.4.3 Reactions to ALPAC Report in the United States

Probably the most thoroughgoing commentary of the ALPAC report in the United States is contained in a memorandum, the draft manuscript of which was made available to me by Zbigniew L. Pankowicz, of the Griffiss Air Force Base, Rome, New York. In it he challenges the committee's report on the following grounds:

... (1) inferior analytical work resulting in factual inaccuracies; (2) hostile and vindictive attitude toward machine translation; (3) use of obsolete and invalid facts and figures as a basis for condemnation of machine translation; (4) distortion of quality, speed and cost estimates in favor of human translation; (5) concealment of data reflecting credit on machine translation *(suppressio veri suggetio falsi),* and (6) wilful omission of dissenting statements on machine translation, presented to the Committee by some experts in this field ([*97*], p.1).

Pankowicz states very carefully that the views expressed in this memorandum are his own and do not represent the opinion of the Rome Air Development Center, at which he is employed. He marshalls

[28] Translation from German is by the author.
[29] Translation from Russian is by the author.

an array of documented evidence which controverts the accuracy of some of the data of the committee, points out some contradictions and omission of some of the evidence gathered by the committee and finds that, “It is obvious that machine translation has been condemned by the committee prior to a sufficient and full examination of the case” ([*97*], p.11) thus echoing in general the feelings of other critics cited above.

I should like to mention just one more commentary[30] on the ALPAC report in the United States by someone not in any way connected with MT. James P. Titus reports arguments against two conclusions of the committee by a group of government research administrators who met in Washington 2 months after the ALPAC report was released. They asserted that there was no surplus of translators, as claimed by the ALPAC report, and that, contrary to other findings of the committee, MT was produced at five facilities around the world. He then asks:

> Why should two intelligent groups of men, all seated in chairs of responsibility, come to such divergent views on the same subject? The main reason seems to be that they are examining different information to form their conclusions. The government research administrators are looking at their projects as they exist today. The ALPAC group looked at data that was probably 2 years old. And that is the basic weakness of the ALPAC report. ([*91*], p.189).

But perhaps the best critique of the ALPAC report is to be found in the observation that the committee not only deprecated the importance of the achievements of MT, but also ignored its future potentialities by implying that MT has no future. Titus ([*91*], p. 191) concludes that

> . . .it seems premature to abandon support of machine translation after only 12 brief years, especially if the abandonment is based only on the findings of the ALPAC committee. Even the critics of machine translation admit its contribution to the knowledge of linguists. Who can say what contributions lie ahead?

He then cites the following comment by a distinguished linguist and serious MT researcher:

> As W.P.Lehmann, Director of the Linguistic Research Center at the University of Texas, put it : “If Dr. Michael E.DeBakey devises a heart pump and it is not immediately successful in its application, the biological community does not raise a great hue-and-cry and returns to theoretical research, shelving the heart pump. It continues experimentation.” ([*91*], p.191).

6. Conclusion

6.1 Justification for Continuing MT

It now becomes clear from the foregoing that even if one admits that FAHQT is not realizable, there is still abundant justification for continuing both MT research and efforts to advance the development of functioning MT systems along the lines carried out by the five groups mentioned in Sections 4.3.1-5 of this survey. The primary objective of the latter would be to transmit information, particularly in the area of natural sciences, from one language to another without regard to established rules of grammar in the target language. In Russian, for example, adverbs of quantity which in English would require the plural can be used with a singular verb, so that one can very easily obtain in English a translation of the type "many industrial zones is situated . . .." It goes without saying that one of the major aims of developing a functioning MT system should be to avoid mis-translations, but one should not expect that practical MT should be more rigorous in this respect than human translation. It is obvious, too, that MT research should also be continued since it contributes to the expansion of our knowledge of how language operates, as already pointed out numerous times in preceding discussions. For instance, research in development of automatic recognition procedures on the syntactic and semantic levels is of help not only in sentence analysis but also in constructing rules for generation of sentences. The implication of the latter for practical MT, information storage and retrieval, and language teaching are quite obvious. Last, MT, generally speaking, is the most important area where symbiosis between computer and man can be most successfully advanced.

6.2 Practical MT as an Application of Symbiosis between Man and Machine

This interaction between man and machine can and should also be applied in the development of a practical MT system. This co-operation should not be limited to the areas of post-editorship of machine produced texts alone. Even after automatic print readers have been developed, certain areas of the input process, like resolving of textural ambiguities will be best left to man, who can handle them more efficiently than the machine can. In the area of dictionary look-up, words of incoming text not found in the glossary stored in the computer, should be translated and grammar coded by man, who again can do it more efficiently than the computer by using his own built-in enormous storage and retrieval system. After the machine stored dictionary has been thus updated, the chances of fouling up the developed syntactic routines are considerably reduced, since even one missing lexical item in the sentence can render the parsing routines inoperative. It is also quite obvious that in the case of failure of resolution of contextual ambiguities by the computer, when two or more target language equivalents for source language lexical items are printed out the man can easily cross out those translations which do not apply. The computer has failed here, but in some cases the resolution of this difficulty by man can be incorporated into the MT system.

6.3 Development of a Guide to Read "MTese"

Since it is quite obvious that the product of any functioning MT system will never look the same as human translation man will also have to expend some effort in reading the output. This process will, however, require less effort than the acquisition of reading skill of another language, which is generally recognised as the easiest to achieve of the three areas of learning a second language, reading, writing, and speaking. Consequently, MT output will have to be accompanied by a brief guide how to read the translation. In essence, these instructions will be statements detailing those structural differences between the source and target languages, obviously from the point of view of mapping the source language into the target language, which have defied automatic analysis at the time of the MT run. These directions will, of course, change as the MT system advances. The compilation of these reading instructions again will be performed by man with the help of the computer, which will indicate to him where these statements are needed.

6.4 MT Most Effective in Small Subarea of Language.

One more element in the implementation of functioning MT systems must be mentioned, the carrying out of the translation activities in a small subarea of language, usually a subfield in the natural sciences, e.g., solid-state physics, marine biology, partial differential equations in mathematics, and others. The smaller the area of language, the greater will be the success in practical MT. One has only to cite the eminently solid results achieved in the Soviet Union in automatically translating American patent literature into English mentioned in Section 4. For this purpose microgrammars and microglossaries have already been and will be created. Appropriately tagged, these computer-stored micrograммars and microglossaries can be modified, expanded, and merged whenever necessary, so that they can be available to all needing them. The computer is, of course, eminently suited to aid substantially in this activity.

To sum up, concentration on small subfields of language expression, creation of guides for reading "MTese" and appropriate division of labor between man and machine are all necessary ingredients for the successful operation of a functioning MT system. Since computers will not go away and are, quite obviously, here to stay, it makes no sense to renounce their application in such an important area of human behavior as language output. You don't throw away an important tool: you use it.

SELECTED BIBLIOGRAPHY[31]

There has been no attempt here to compile an exhaustive bibliography of documents related to the field of automatic language processing. The references contained herein were selected primarily for the purpose of providing a source of general information on, and broad coverage of, research in mechanical translation. Readers who are interested in the particular work and publications of individual MT research groups described in this survey should address their inquiries directly to the pertinent groups.

PROCEEDINGS AND CONFERENCES

Abstracts of papers presented at the Second (1964), Fourth (1966), Fifth (1967), Sixth (1968) Annual Meetings of the Association for Computational Linguis-
tics (formerly the Association for Machine Translation and Computational Linguistics) available at the Slavic Department, Wayne State University, Detroit, Michigan 48202.

Edmundson, H.P. ed., *Proc. Nat. Symp. Machine Translation, Los Angeles, 1960. Prentice Hall, Englewood* Cliffs, New Jersey 1961.

1965 International Conference on Computational Linguistics, New York, 1965. Reprints available at the Slavic Department, Wayne State University, Detroit, Michigan 48202.

Proceedings of the 1961 International Conference on Machine Translation of Languages and Applied Language Analysis, National Physical Laboratory, Teddington, England 1961, Vol. I-II. London, H.M. Stationary Office 1962.

Proceedings of the Conference on Computer-Related Semantic Analysis. Machine Translation Research Group, Wayne State University, Detroit, Michigan, December 1965.

Second International Conference on Computational Linguistics (2ème Conference Internationale sur le Traitement Automatique des Langues), Grenoble, France, August 23-25, 1967.

Summary of the Proceedings of the Wayne State University Conference of Federally Sponsored Machine Translation Workers. Machine Translation Research Group, Wayne State University, Detroit, Michigan, July 1900.

[31] See also references contained therein.

Summary of the Proceedings of the Russian-to-English Grammar Coding Conference. Machine Translation Research Group, Wayne State University, Detroit, Michigan, April 1961.

Summary of the Proceedings of the Conference of Federally Sponsored Machine Translation Groups on MT-Oriented Syntactic Analysis. Machine Translation Group, Wayne State University, Detroit, Michigan, June 1962.

PERIODICALS

Computational Linguistics. Computing Centre of the Hungarian Academy of Sciences, Budapest, Hungary.

The Finite String. Newsletter of the Association for Computational Linguistics (formerly The Association for Machine Translation and Computational Linguistics). A. Hood Roberts, ed., Associate Director, Center for Applied Linguistics, 1717 Massachusetts Ave. N.W., Washington, D.C. 20036. (Each issue includes a list of recent publications in computational linguistics and. machine translation.)

Foreign developments in machine translation and information processing. U.S. Department of Commerce, Office of Technica1 Services, Joint Publications Research Service (JPRS), Washington, D.C. 20025.

KVAL. Research Group for Quantitative Linguistics. Fack, Stockholm 40, Sweden.

Mechanical Translation. Published at the University of Chicago Press for the Association for Computational Linguistics. Victor Yngve, ed., Graduate Library School, University of Chicago, Chicago, Illinois 60637.

The Prague Bulletin of Mathematical Linguistics. Karlová Universita, Prague, Czechoslovakia.

Prague Studies in Mathematical Linguistics. Academia, Czechoslovak Academy of Sciences, Prague, Czechoslovakia..

Revue Roumaine de Linguistique. L'Académie de la Republique Socialiste de Roumanie, Bucharest, Rumania.

Statistical Methods in Linguistics (SMIL). Skriptor, Fack, Stockholm, 40. Sweden.

La Traduction Automatique, Bulletin trimestriel de 1'Association pour l'étude et le Développement de la Traduction automatique et de la Linguistique appliquée (ATALA), 20, rue de la Baume, Paris 8e, France.

Two leading Soviet references *Научно-техническая Информация* (Scientific Technical Information), Academy of Sciences, USSR , Moscow. Issues starting with 1968 include tables of contents and summaries of each article in English translation.) *Реферативный Журнал* (Abstract Journal), Academy of Sciences, USSR, Moscow. This journal also appears completely in English translation.) See especially the section headed Автоматический перевод текстов (Automatic Text Translation)

GENERAL REFERENCES

Akhmanova, O. S., Mel'chuk, I. A., Frumkina, R. M., and Paducheva, E.V. *Exact Methods in Linguistic Research.* Hays, D. G., and Mohr, D. V., The RAND Corporation. Transl. Santa Monica, California, September 1963.

Booth, A. D., Brandwood, L., and Cleave, S. C. *Mechanical Resolution of Linguistic Problems.* Academic Press. New York, and Butterworths, London 1958.

Delavenay, E. *An Introduction to Machine Translation.* Thames and Hudson, London. 1960.

Garvin, P. L. and Spolsky, B., eds. *Computation in Linguistics, A Case Book.* Indiana Univ. Press, Bloomington, Indiana, 1966.

Hays, D. G., Research Procedures in Machine Translation. RAND Research Memo. RM-2916-PR, December 1961.

Hays, D. G., ed., *Readings in Automatic Langauge Processing.* American Elsevier, New York, 1966.

Hays, D., G., *Introduction to Computational Linguistics.* American Elsevier, New York, 1967.

Language and Machines (The ALPAC Report). Computers in Translation and Linguistics. Automatic Language Processing Advisory Committee, Report. Publication 1416, Behavioral Sciences, Nat. Acad. Sci. Nat. Res. Council, Washington, D.C., 1966.

Lehmann, W. P., *Machine Translation Research During the Past Two Years,* p. 10. Erie Document 020503, 1968. Information concerning ERIC publications is available from ERIC Document Reproduction Service, 4936 Fairmont Ave., Bethesda, Maryland.

Locke, W. N., and Booth, A. D., eds., *Machine Translation of Languages.* M.I.T. Press, Cambridge, Massachusetts, 3rd printing, 1965.

Машиный перевод Трыды Института точной механики и вычислительной техники АП СССР, Москва, вып. 1, 1958г; вып. 2, 1961г. [*Machine Translation.* Institute for Precision Mechanics and Computational Technology of the Academy of Sciences of the U.S.S.R., Moscow, Issue 1, 1958; Issue 2. 1961.]

Машиный перевод и прикладная лингвистика. Труды московского государственного педагогического инстиута иностранных язков, Москва, вып.8, 1964. [*Machine Translation and Applied Linguistics.* Moscow State Foreign Language Pedagogical Institute, Moscow, Issue 8, 1964.

Мельчук, И.А. *Автоматический синтаксический анализ*. Редакционно-издательский одел сибирского отделения АН СССР, Новосибирск, 1964. [Mel'chuk, I. A., *Automatic Syntactic Analysis.* Department of the Siberian Branch of the Academy of Sciences of the U.S.S.R., Novosibirsk, 1964.

Мельчук, И.А и Равич, Р.Д. *Авоматический перевод*, 1949-1963. Винити, Институт язкознания, москва, 1967. [Mel'chuk, I. A. and Ravich, R. D., *Automatic 'Translation,* 1949 - 1963. All Union Institute for Scientific and Technical Information. Linguistics Institute, Moscow, 1967.

Nida, E. A., *Toward a Science of Translating. E.* J. Brill, Leiden 1964, see, especially the chapter entitled Machine translation.

Oettinger, A. G. *Automatic Language Translation.* Harvard University Press, Cambridge, Massachusetts, 1960.

Papp, F. *Mathematical Linguistics and Mechanical Translation in the Soviet Union.* Mouton, The Hague, The Netherlands. November 1965.

Ревзин, И.И. и Розенцвейг, В.Ю., *Оснобы общего и машинного перевода*, изд-но «Высшая Школа», Москва, 1964. [Revzin, I. I. and Rozentsveyg, V. Yu., *Principles of General and Machine Translation.* Higher Education Publishing House, Moscow, 1964.]

Vauquois, B. A survey of formal grammars and algorithms for recognition and transformation in mechanical translation. IFIP Congress, 1968 Preprints, North-Holland Publ., Amsterdam, 1968.

Yngve, V., Implications of mechanical translation research. *Proc. Am. Phil. Soc.* **108**, No. 4, 275 (1964).

REFERENCES

1. Bar-Hillel, Y., The present status in automatic translation of languages. *Advan.Computers* **1**. 1960-1961 (1960).
2. Whorf, B. L.. The relation of habitual thought to language in *Language, Culture and Personality* (L. Spier, ed.), pp. 75-93. Sapir Memorial Publ. Fund, Menasha, Wisconsin. 1941.
3. Josselson, H. H., Automatization of lexicography. *Cah. Lexicol.* **9**, 73-87 (1966).
4. Josselson, H. H., Lexicography and the computer, in *To Honor Roman Jacobson,* pp. 1046-1059. Mouton, 's-Gravenhage, The Netherlands, 1967.
5. Booth, A. D., and Locke, W. N., Historical introduction, in *Machine Translation of Languages, pp.* 1-14, M. I. T. Press, Cambridge, Massachusetts, 3rd Printing, 1965.
6. Kuznetsov, V. I., Preamble *Proc. Symp. Machine Translation Budapest, October 1967.* see *Nauch. Tekh. Inform.* (Scientific-Technical Information) 2, No. 6, 28-30 (1968).
7. *Current Research and Development in Scientific Documentation,* **11.** Nat. Sci. Found. Washington, D.C. 1962.
5. Edmundson. H.P., ed., *Proc. Nat. Symp. Machine Translation, Los Angeles, 1960.* Prentice-Hall, Englewood Cliffs, New Jersey, 1961.
9. Garvin, P. L.. An informal survey of modern linguistics. *Amer. Doc.* **10,** 291-298 (1965).
10. Models and Theories Session, *Proc. Conf. Federally Sponsored Machine Translation Groups MT-Oriented Syntactic Analysis. Princeton, 1962,* pp. 83-92. MT Research Dept., Wayne State University, Detroit, Michigan, 1962.
11. Brower, R. A., ed., *On Translation.* Harvard Univ. Press, Cambridge, Massachusetts, 1959.
12. Chao, Y. R., Translation without machine, pre-publication copy of paper delivered at the IX International Congress of Linguists, August 27-31. Cambridge, Massachusetts, p. 6.
13. Sherry, M., The identification of nested structures in predictive syntactic analysis. *Proc. 1st Intern. Conf. Machine Transl. Lang Appl. Lang. Anal., Teddington, England, 1961.* pp.1-13. H.M. Stationery Office, London 1962.
14. Alt, F., Recognition of clauses and phrases in machine translation of languages. Nat. Bur. Stds. Rept. 6895, Washington. D.C. 1960.
15. Yngve, V., A model and an hypothesis for language. *Proc. Amer. Phil.Soc.,* **104,** 444-446 (1960).
16. Sager, N., Procedure for left-to-right recognition of sentence structure. Transformations and Discourse Analysis Proj. Rept 27. University of Pennsylvania, Philadelphia, Pennsylvania, 1960.
17. *Proc. Conf. Federally Sponsored Machine Translation Groups MT-Oriented Syntactic Analysis. Princeton, 1962,* pp. 79-82. Wayne State University, Detroit. Michigan, 1962.
18. Hays, D. G. Dependency theory: A formalism and some observations. RAND Memo RM 4087-PR. July 1964; for an earlier discussion of a related topic, see Tesnière, L. *Esquisse d'une Syntaxe Structurale*. Klincksieck, Paris, 1953.
10. Lamb, S. M., *Outline of Stratificational Grammar.* Georgetown University Press, Washington, D.C., 1966.

20. Bach, E., *An Introduction to Transformational Grammars,* Holt, New York, 1964.
21. Senechalle, D., Introduction to formation structures, Rept. LRC63 WTM-2. University of Texas, Austin, Linguistics Research Center, April 1963.
22. Estes, W. B., Holley, W, H., and Pendergraft, E.D., Formation and transformation structures, Rept. LRC63 WTM-3. University of Texas, Austin, Linguistics Research Center, May 1963.
23. Ceccato, S. ed., *Linguistic Analysis and Programming for Mechanical Translation, G*ordon and Breach, New York, 1961.
24. Dutton, B., An introduction to the theory and practice of correlational grammar. Georgia, Institute for Research, Athens, Georgia.. October 1968.
25. Rhodes, I., A new approach to the mechanical translation of Russian, Nat. Bur. Stds. Rept. 6295. Washington, D.C., February 1959.
26. Sherry, M., Mathematical linguistics and automatic translation, Nat. Sci. Found. Rept. 5. Harvard University, Cambridge, Massachusetts, August 1960.
27. Plath, W., Mathematical linguistics and automatic translation, Nat. Sci. Found. Rept. 12. Harvard University, Cambridge, Massachusetts, June 1963.
28. Machine translation studies of semantic techniques, RADC-TR 61-72. Ramo-Wooldridge, Canoga Park, California, February l961.
29. Studies in distributional semantics, in Six tasks in computational linguistics, Chapter 2, RM-280-AFOSR. RAND Corporation, Santa Monica, California, October 1961.
30. Hays, D. G., Research procedures in machine translation, pp. 18-25, 33-36, 40-48. RM-2916-PR. RAND Corporation, Santa Monica, California, December 1961.
31. Householder, F. W., Jr., and Lyons, .J, *Fourth Quart. Rept. Automation Gen. Semantics,* pp. 8-9. Indiana University, Bloomington, Indiana, February 1961.
32. *Proc. Conf. Computer-Related Semantic Analysis, Las Vegas, Nevada, 1965,* held under the auspices of Wayne State University.
33. Linguistic and engineering studies in automatic language translation of scientific Russian into English, Tech. Rept, RADC-TN-58-321. ASTIA Doc. AD-148992. Dept. Far Eastern and Slavic, Languages, and Dept. Electrical Engineering, Univ. of Washington, Seattle, Washington, June 1958.
34. Translation study, Final Rept. AF30(602)-2131, RADC-TR-61-235. Dept. Electrical Engineering, Univ. of Washington, Seattle, Washington, August 28, 1961.
35. The Chinese-English machine translation project, Vols. I and II. Final Rept. Dept. Far Eastern and Slavic Languages and Literatures, Univ. of Washington, Seattle, Washington, September 1962.
36. Yngve, V., *MT at M.I.T., 1965.* Research Laboratory of Electronics, Mass. Inst. Technol., Cambridge, Massachusetts, June, 15, 1965.
37. Macdonald, R. R., General report 1952- 1963, Occasional papers on machine translation, No. 30, Georgetown Univ. Machine Translation Res. Proj. Georgetown University, Washington, D.C., June 1963.
38. Hays, D. G., Research procedures in machine translation, Memo. RM-2916-PR. The RAND Corporation, Santa Monica, California, December 1961.
39. Harper, K. E., Hays, D.G., and Scott, B.J., Studies in machine translation bibliography of Russian scientific articles, Memo. RM-3610-PR. The RAND Corporation, Santa Monica, California, June 1963. (This publication includes a list of RAND studies in machine translation on page 97.)

40. Hays, D. G.. Annotated bibliography of RAND publications in computational linguistics, Memo. RM-3894-PR. The RAND Corporation, Santa Monica, California., March 1964; see also Computational Linguistics Bibliography, 1967. Memo. RM-5733-PR, June 1968.
41. Hays, D. G. Dependency theory: A formalism and some observations. Memo. RM-4087-PR. The RAND Corporation, Santa Monica, California, July 1964.
42. Machine translation studies of semantic techniques, Final Rept. AF30(602)-2030. Ramo-Wooldridge Div., Thompson-Ramo-Wooldridge, Canoga Park, California, February 22, 1961.
43. Fulcrum techniques to semantic analysis, Final Rept. AF30(602)-2643. TRW Computer Div., Thompson-Ramo-Wooldridge, Canoga Park, California, March 5, 1963.
44. A syntactic analyzer study, Final Rept. AK30(602)-3506. Bunker-Ramo Corp., Canoga Park, California., July 31, 1965.
45. Adaptation of advance fulcrum techniques to MT production system (Russian-English), AF30(602)-3770. Bunker-Ramo Corp., Canoga Park, California, November 1, 1966.
46. Computer-aided research in machine translation. Final Rept. NSF-C372. Bunker-Ramo Corp., Canoga Park, California, March 31, 1967.
47. Oettinger, A., *Automatic Language Translation,* Harvard Monographs Appl. Sci., No. 9. Harvard Univ. Press, Cambridge, Massachusetts. 1960.
48. Sherry, M., Comprehensive report, on predictive syntactic analysis, in Mathematical Linguistics and Automatic Translation, Rept. No. .NSF-7, Section I. Harvard Computation Laboratory, Cambridge, Massachusetts, 1961.
49. Oettinger, A, G., Automatic syntactic analysis and the pushdown store. *Proc. Symp. Appl. Math.* **12.** Am. Math. Soc., Providence, Rhode Island, 1961.
50. Mathematical linguistics and automatic translation, Rept. No. NSF-13. Harvard Computation Laboratory, Cambridge, Massachusetts, March 1964.
51. Mathematical linguistics and automatic translation, Rept. No. NSF-21. The Harvard Russian Syntactic Analyzer Manual. Harvard Computation Laboratory, Cambridge, Massachusetts, May 1968.
52. Koutsoudas, A., Machine translation at the University of Michigan, in *Information Retrieval and Machine Translation,* Part II (Allen Kent, ed.), pp. 765-770. Interscience, New York, 1961.
53. Smoke, W., and Dubinsky, E., A program for the machine translation of natural languages. *Mech. Transl.* **6**, 2-10 (1961).
54. Rhodes, I., A new approach to the mechanical translation of Russian, NBS Rept. 6295. U.S. Department of Commerce, Nat. Bur. Stds., Washington. D.C., February 6, 1959; see also, A new approach to the mechanical syntactic analysis of Russian, NBS Rept. 6595, November 10, 1959.
55. Alt, F., Recognition of clauses and phrases in machine translation of languages. NBS Rept. 6895. U.S. Department of Commerce, Nat. Bur. Stds., Washington. D.C., July 11, 1960.
56. Research in machine translation, Ann. Repts. 3, 4, and 5. Wayne State University, Detroit, Michigan, 1961, 1962, and 1963.
57. Steiger, A. J. Parsing by matrix. MT Research Detroit, Wayne State University, Detroit, Michigan, 1965.
58. Ten year summary report, MT Research Dept., Wayne State University, Detroit, Michigan, May 1968.

59 Lamb, S. M., and Jacobson, Jr. W. H., A high-speed large-capacity dictionary system. *Mech. Transl.* **6,** 76-107 (1961).

60. Lamb, S. M., On the mechanization of syntactic analysis. *Proc. 1st Intern. Conf. Machine Transl. Lang. Appl. Lang. Anal., Teddington, England, 1961.* paper 21. H.M. Stationery Office, London, 1962.

61. Lamb, S. M., Outline of stratificational grammar. University of California, Berkeley, California, 1962.

62. Final report on machine translation study. Rept. No. 16, Linguistics Research Center, University of Texas, Austin, Texas, June 1963.

63. Tosh, W., *Syntactic Translation.* Mouton 's-Gravenhage, The Netherlands, 1965.

64. Research on syntactic and semantic analysis for mechanical translation, Final Rept., Linguistic Research Center, University of Texas, Austin, Texas, April 1967.

65. Levison, M., The Mechanical analysis of language. *Proc. 1st Intern. Conf. Machine Translation Lang. Appl.Lang. Anal., Teddington, England, 1961,* paper 29. H.M.Stationery Office London, l962.

66. Parker-Rhodes, A. F. A new model of syntactic description. *Proc. 1st Intern. Conf. Machine Translation Lang. Appl. Lang. Analysis, Teddington, England, 1961,* paper 34. H.M. Stationery Office, London. 1962.

67. Masterman, Margaret. Semantic algorithms. *Proc. Conf. Computer-Related Semantic Analysis,* Machine Translation Res. Proj., Wayne State University, Detroit, Michigan, 1965. See also Karen Spark-Jones, Semantic classes and semantic message forms, same conference.

68. Linguistic problems of scientific information, ICSU-UNESCO, Joint Project on the Communication of Scientific Information. Moscow, November 19, 1968.

69. Ceccato, S., and Zonta, B., Human translation and translation by machine. *Proc. 1st Intern. Conf. Machine Translation Lang. Appl. Lang. Anal., Teddington, England, 1961, pap*er 30. H.M. Stationery Office, London 1962.

70. Ceccato, S.. Automatic translation of languages. *Inform. Stor. Retrieval* **2**, 105-158 (1965).

71. Ceccato, S., Correlational analysis and MT, in *Progress in Mechanical Translation.* North Holland Publ., Amsterdam, The Netherlands, 1966.

72. Nedobejkine Maksimenko, N., and Torre, L., Modèle de la Syntaxe Russe, Centre d'Etudes pour la Traduction Automatique, Grenoble, France. December 1964.

73. Rapport d'Activite pour l'Année 1966, Centre d'Etudes pour la Traduction Automatique, Grenoble, France, 1966.

74. Baille. A., and Roualt, J., Un Essai de Formalisation de la Semantique des Langues Naturelles. Centre d'Etudes pour la Traduction Automatique, Grenoble, France, December 1966.

75. Veillon, G., Veyrunes, J., and Vauquois, B., Un Metalangage de Grammaires Transformationelles, Centre d'Etudes pour la Traduction Automat ique, Grenoble, France, 1967.

76. Querido, A., Deux Modèles de Description Syntaxique, Centre d'etudes pour le traitement automatiques des données linguistiques, University of Montreal, July 1966.

77. Friant, J., Grammmaires Ordonées — Grammaires matricielles, Centre d'études pour le traitement automatique des données linguistiques. University of Montreal, October 1968; Langages Ultralinéaires et Superlinéaires Nouvelles Caracterisations, Centre d'études pour traitement automatique des données linguistiques, University of Montreal, November 1968.

78. Hirschberg, L. Dictionnaires Automatiques pour Traducteurs Humains. *J. Traducteurs,* Brussels, 1966.
79. Konečna, D., Novák, P., and Sgall, P., Machine translation in Prague. *Prague Studies in Mathematical Linguistics.* No. 1. Czechoslovak Academy of Sciences, Prague. Czechoslovakia, 1966.
80. Varga, D., Towards a new system of automatic analysis. *Computat. Linguistics* **6**, l23-136 (1967).
81. Kunze, J., Theoretical problems of automatic translation. Z. *Mat. Logik Grundlagen der Matematik.* **12**, 85-130 (1966).
82. Nishimura, H., The YAMATO and the experimental translation, 46 pp. Electrotechnical Laboratory, Tokyo, Japan, May 1964.
83. Sakai, I., Some mathematical aspects of syntactic description, paper presented at the International Conference of Computational Linguistics, New York, May 1965.
81. Sakai, T., Models and strategies for MT, presented at the U.S.-Japan Seminar on MT. Tokyo, Japan, May l964.
85. Tamati, T., Syntactic description of Japanese grammar, presented at the U.S.-Japan Seminar on Mathematical Translation, New York, 1965.
86. Brown, A., *The "SLC" Programming Language and System for Machine Translation.* Vol. I: *Programming Manual* (rev. ed.); Vol. 2; *Utility Programs.* Rept. No. EUR2418, Euratom, Brussels, Belgium, 1964.
87. *An Evaluation of Machine-Aided Translation Activities at FTD,* Arthur D. Little. May 1, 1965.
88. McDaniel, J., *et al.,* Translation of Russian Scientific Texts into English by Computer. Final Rept., Autonomics Div., National Physical Laboratory, Teddington, England, July 1967.
89. Kravec, L.G., Vasilevskij, A. L., and Dubickaja., A. M., Eksperimental'naja sistema avtomaticheskogo perevoda publickacij iz Amerikanskogo patentnogo ezhenedel'nika "Official Gazette." *Nauch.-Tekh. Inform.* Ser. **2**, .No. 1, 35-40 (1967).
90. Language and machines, computers in translation and linguistics, Publication 1416, pp. 107 - 112. Automatic Language Processing Advisory Committee Report, (ALPAC Report), Nat. Acad. Sci. Nat. Res. Council, Washington, D.C.
91. Titus. J. P.. The nebulous future, of machine translation. *Commun. Assoc. Comput. Machinery* **10**, No. 3, 190 (1967).
92. *The Finite String,* **3,** October/November 1966.
93. Krollman, F., *Der Sprachmittler,* No. 3, 85 (l968).
94. Krollman, F., *Der Sprachmittler.* No. 4, 121 (1967).
95. Krollman, F., *Der Sprachmittler,* No. 2, 43 (1968).
96. Nauchno-Tekhnicheskaja Informatsija. Scientific Technical Information, Ser. **2**. No. 8, 24 (1968).
97. Pankowicz., Z.L., Draft of a Commentary on ALPAC Report, Part I. Rome Air Development Center, Griffiss Air Force Base, Rome, New York, March 1967.
98. Garvin, P. L., Machine translation—fact or fancy? *Datamation* pp. 29-31, 1967.