FEASIBILITY STUDY ON FULLY AUTOMATIC HIGH QUALITY TRANSLATION

Dr. Winfred P. Lehmann
Dr. Rolf Stachowitz

University of Texas

FOREWORD

This is the Final Report, in two volumes, for the Feasibility Study on Fully Automatic High Quality Translation, by the University of Texas, Linguistic Research Center, Austin, Texas, for Rome Air Development Center, Griffiss Air Force Base, New York, under contract F30602-70-C-0129, Job Order No. 45940000. Zbigniew L. Pankowicz (IRDT) was the RADC Project Engineer.

As the appendices indicate, the study brought together specialists in the areas involved in machine translation. The report summarizes their findings. Participants in the study were provided with a preliminary statement of the initial part of this report, except for the conclusions and recommendations, and were asked to send their comments and revisions. These were incorporated in this report, except when they did not seem in keeping with the general conclusions of the various other participants. There were few strikingly diverse points of view.

The authors are grateful to the participants, and to the sponsors who made the study possible .

This report has been reviewed by the Information Office (OI) and is releasable to the National Technical Information Service (NTIS) .

This technical report has been reviewed and is approved.

Approved: Zbigniew L. Pankowicz

ZBIGNIEW L. PANKOWICZ
Technical Evaluator

Approved: F. H. Dettmer

FRANZ H. DETTMER
Colonel, USAF
Chief, Intel and Recon Division

FOR THE COMMANDER: Fred I. Diamond

FRED I DIAMOND
Acting Chief, Plans Office

ABSTRACT

This report presents the results of a theoretical inquiry into the feasibility of a fully automatic high quality translation (FAHQT), according to Bar-Hillel's definition of this term. The purpose of this inquiry consisted in determining the viability of the FAHQT concept in the light of previous and projected advances in linguistic theory and software/hardware capabilities. The corollary purpose was to determine whether this concept can be taken into consideration as a legitimate and justifiable objective of R&D. The effort was supported by 20 expert consultants from the various universities and research centers in the U.S.A. and abroad. Conclusions and recommendations are presented on pages 44-50 of the report. Individual contributions of participants and consultants reflect a wide range of opinions concerning the prospects of FAHQT in intermediate and long range of R&D.

TECHNICAL EVALUATION

The objective of this theoretical inquiry is to examine the controversial issue of a fully automatic high quality translation (FAHQT) in the light of the past and projected advances in linguistic theory and hardware/software capability. The principal purpose of this study is to determine whether the concept of FAHQT is justifiable as a long range R&D proposition. The study is also concerned with the intermediate range alternatives to FAHQT, i.e., machine translation forms that are adequate to the user's needs with or without post-editing. Machine aided translation, based on the automated dictionary look-up, is excluded from the study in consideration of the fact that this by-product of machine translation R&D is well within the current state-of-the-art.

In the context of FAHQT, "full automation" implies that the entire translation process is autonomous in the computer without pre-editing of the source language text and post-editing of the target language output. "High quality" seems to be undefinable in an absolute sense. In referring to machine translation of 100% quality, Bar-Hillel (1) introduced the following qualification.

> "When I talk about "100%", I obviously have in mind not some heavenly ideal of perfection, but the end product of an average human translator. I am aware that such translator will on occasion make mistakes and that even machines of a general low quality output will avoid some of these mistakes. I am naturally comparing averages only".

Thus viewed, even the concept of 100% quality is not equatable with the error-free performance in either form of translation. Understandably enough, participants and consultants failed to reach a unanimous agreement as to the definition of "high quality" in machine translation. This is reflected on p. 48, quote, "There is apparently no absolute standard. Rather, standards must be defined with reference to specific users and specific purposes". In the absence of absolute and universally valid quality criteria, the user of machine translation can be legitimately considered an ultimate judge of its quality. This viewpoint was first expressed by Reitwiesner and Weik (2) as early as in 1958.

According to Lamb (3), "all translation can be viewed as human translation since machine translation is nothing but another kind of human translation". It follows from this observation that the fundamental constraints on machine translation parallel those imposed on human translation. Assuming the well-known limits of translatabi1ity, this seems to imply that either form of translation is a priori constrained. In summarizing the problem of translation equivalence between SL (source language) and TL (target language),

Catford (4) draws the following conclusion.

> "The limits of translatabi1ity in total translation are, however, much more difficult to state. Indeed, translatability here appears, intuitively, to be a cline rather than a clear-cut dichotomy. SL texts and items are more or less translatable rather than absolutely translatable or untranslatable. In total translation, translation equivalence depends on the interchangeabi1ity of the SL and TL texts to (at least some of) the relevant features of situation-substance".

Ray (5) recognizes the fact that "every translation necessarily involves some distortion of meaning". However, as is reflected in his statements below, this deficiency is not only manageable, but even unimportant in the practice of translation.

> "The translation operation is, like the limit operation, possible only under such conditions as "sufficiently" and "arbitrarily", that is, only by the exercise of some evaluative judgement, however little. Since distortion of meaning cannot be avoided, the problem becomes one of confining it to allowable measures of allowable kinds in allowable places along allowable directions".
>
> "..., while no two languages will match exactly in the total range of possible discourse, there are infinitely many specific limited ranges of discourse where the distortion of meaning can be legitimately dismissed as of no account".

The feasibility of FAHQT must be, therefore, considered within the limits of translatability, i.e., taking into account the constraints on the total-translation. Since the concept of high quality is untenable in the absolute sense, the question of what is feasible in the context of FAHQT is quite probably more meaningful. It would be patently unreasonable in this stage of R&D to postulate machine translation requirements beyond the limits of translatabi1ity imposed on human translation.

Machine translation research, based on puristic notions and oriented toward a global solution, was once compared to a search for the Holy Grail. This all-or-nothing attitude has probably caused as much damage to the progress of machine translation research as the early announcements of quick and easy solutions. Perfectionists in this area have generally tended to ignore the injunction by Lecerf (6) that "entreprendre la mise au point d'ensembles de traduction automatique, c'est avant tout accepter la contrainte du reel".

According to Ljudskanov (7),

> "The widespread so-called 100 percent approach, along with the belief that MT presupposes the presence of a complete mathematical model of language in general and of the specific languages in particular, in practice amounts to equating the nature and extent of the knowledge of language in general, which is necessary from the point of view of theoretical linguistics, with the extent of knowledge necessary for the achievement of translation from one language into another. This approach also amounts to equating the description of communication in general with that of the translation process; it ignores the specific characteristics of the process as mentioned above and the general linguistic problems of the theory of translation (both HT and MT) in the general problem area of mathematical linguistics".

> "....it can be asserted that the current critical state of MT research throughout the world, although much has happened that legitimately causes well-grounded anxieties and doubts as to its possibilities, is due to a certain degree to the maximalistic tendencies, however laudable they may be in themselves, of the global strategy. By giving due consideration to the particular characteristics of the translation process and of its study, as well as to the differentiation of the aims of mathematical linguistics from the theory of MT and of the fields of competence and performance from each other, research in this field would be channeled in a direction both more realistic for our time and more closely in accord with the facts".

The report highlights on p.4 an important, but often ignored, difference between scientific and technical translations and translations of literary and religious texts, in spite of its importance from the viewpoint of machine translation requirements.

> "Even articles and monographs dealing with machine translation have failed to be adequately explicit about the special problems of translating technical and scientific materials by computer. Instead, they have confused the problem by comparing machine translation with the long-practiced human translation, by equating the problems of translating scientific materials with those involved in translating literary materials, and by using the same evaluation criteria for the results".

It is now a commonplace that the style of writing is of paramount importance in literary translation, whereas the accuracy constitutes the most important quality criterion in scientific and technical translations. According to Gingold (8),

> "It is not the translator's job to abstract, paraphrase, or improve upon the author's statements. He cannot be expected to convert an article that is poorly organized and badly written in the original language into a masterpiece of English scientific writing. In technical translation, he must always be willing to sacrifice style on the altar of accuracy".

Savory (9) has expressed a similar opinion in his statement that "the translation of scientific work is an ideal example of translation of a writing in which the subject matter is wholly on the ascendant and the style is scarcely considered".

The report further emphasizes the crucial importance of timeliness in production of scientific and technical translations. According to the statement on p. 5, "...timeliness is of increasing importance to users of scientific translations. Even in a relatively unhurried field like linguistics, few articles retain their importance over a long period. Statements have been made repeatedly about the obsolescence of publications issued a few years earlier. The insistence among technical specialists and scientists for speedy translation contrasts markedly with the length of time permitted for completing literary translations". The requirement of timeliness was stressed elsewhere by Gingold (10), quote, "The delay between the appearance of the original journal and its English translation, which may be a year or more, is also a disadvantage, particularly to industry, where time is usually of great importance".

The principal findings of the study, as related to its objectives, can be summarized as follows.

Computer hardware is no longer considered a crucial problem in machine translation. "Remarkable improvements, especially in rapid-access storage devices, have largely eliminated the problems caused by inadequate computers. Lexical items can now be retrieved as rapidly as were the major syntactic rules a decade ago. And with further improvements of storage devices in process, computers no longer pose major problems in machine translation". (p. 12). Developmental prospects in this area are very bright indeed, particularly with the advent of holographic memories. The impact of such memories on both linguistic and computational aspects of machine translation R&D is discussed in detail by Stachowitz in one of his contributions to the report ("Requirements for Machine Translation: Problems, Solutions, Prospects", pp 409-532). This contribution is considered significant because it provides a complete blueprint for a realistic implementation of a large-scale machine translation system.

Equally encouraging is the appreciation of advances in computer software. "Programming has evolved as rapidly as have computers... A key factor here was the enrichment of programming language data types which made possible efficient representation and manipulation of linguistic structures". (p. 13).

The report reflects a unanimous agreement of participants and consultants that "the essential remaining problem is language" (pp 14-15). It is, therefore, not surprising that linguistics has received much more attention in the study than computer hardware and software. Recommendations presented on pp 49-51 are exclusively oriented toward linguistic research in the context of machine translation.

The report points out that there is "no conflict between specialists in descriptive linguistics, linguistic theory and machine translation... As descriptive linguists improve their understanding of language, and the models by which to express that understanding, machine translation specialists will update their procedures and models".(p. 24). However, the report also reflects a difference of opinions between machine translation experts and linguists as regards the nature, orientation and scope of linguistic research involved in machine translation. It is further worth noting that some linguists participating in this study have not acknowledged Ljudskanov's caveat about "maximalistic tendencies of the global strategy".

The reader is referred to Conclusions (pp 45-48) and Recommendations (pp 49-51), summarizing the results achieved in performance of this study. Recommendation of support for research in machine translation is based on the fact that "quality translations can be achieved in the near future. This recommendation agrees strikingly with conclusions reached in a study carried out in the Soviet Union". (p. 49). Galilei's challenge ("Eppur si muove!"), aptly chosen as a motto in the Introduction to (11) by Kulagina and Mel'chuk, would be equally appropriate as an expression of views and sentiments embodied in the main part of this report.

ZBIGNIEW L. PANKOWICZ
Technical Evaluator

BIBLIOGRAPHY:

1. Yehoshua Bar-Hillel, Language and Information, Addison-Wesley Publishing Co., Reading, Mass., 1964 (p. 171).

2. G. W. Reitwiesner and M. H. Weik, Survey of the Field of Mechanical Translation of Languages, Ballistic Research Laboratories Memorandum Report No. 1147, May 1958.

3. Sydney M. Lamb, The Nature of the Machine Translation Problem, Journal of Verbal Learning and Verbal Behavior, 4, 196-210, 1965.

4. J. C. Catford, A Linguistic Theory of Translation, Oxford University Press, 1957 (p. 93).

5. Punya Sloka Ray, A Philosophy of Translation, Babel, Vol. XII, No. 1, 1966 (p. 182).

6. Yves Lecerf, Introduction a l'Ensemble des Quatre Lectures, EUR/C-IS/1234/62 f.

7. A Ljudskanov, Is the Generally Accepted Strategy of Machine Translation Research Optimal?, Mechanical Translation, Volume 11, Number 1 and 2, March and June 1968, p. 19-20.

8. Kurt Gingold, A Guide to Better Translations for Industry, Babel, Vol. XII, No. 3, 1966 (p. 144)

9. Theodore H. Savory, The Art of Translation, London, 1957.

10. Kurt Gingold, Translation in American Industry. Babel, Vol. X, NO. 3, 1966 (p. 115).

11. O. S. Kulagina and I. A. Mel'chuk, Automatic Translation: Some Theoretical Aspects and the Design of a Translation System in Machine Translation (ed. A. D. Booth), John Wiley & Sons, Inc. New York, North Holland Publishing Co., Amsterdam, 1967.

TABLE OF CONTENTS

1. Requirements of translation

With the increase in communication and in publication, translating has achieved a greater importance than ever before. Literary figures are engaged in translating from many exotic languages, as well as from the traditional languages of western culture. Symposia on translation have been held, resulting in the publication of monographs on the topic. Above all, scientists and technical specialists have come to demand translations. As one of the leading experts, Eugene A. Nida, has stated in his most recent contribution to the topic (Nida and Taber, 1969, 1): "Never before in the history of the world have there been so many persons engaged in the translating of both secular and religious materials." The book intimates that the requirements for translation will be increased.

Moreover, it describes more specifically and concretely than earlier discussions the steps that are involved in translation. Translation is defined (Nida and Taber, 1969, 12) as "reproducing in the receptor language the closest natural equivalent of the source-language message. " And the paragraph continues: "this relatively simple statement requires careful evaluation of several seemingly contradictory elements."

For a fuller statement on the problem of translation, we refer to the important books by Nida and their bibliographies. His last book, however, contains further perceptive statements that are important to include here.

A section on "the old focus and the new focus" of translating (Nida and Taber, 1969, 1) states that "the older focus in translating was the form of the message.. .The new focus, however, has shifted from the form of the message

to the response of the receptor."

Further, "even the old question: Is this a correct translation? must be answered in terms of another question, namely: For whom?" After a brief answer, the section continues: "In fact, for the scholar who is himself well acquainted with the original, even the most labored, literal translation will be correct, for he will not misunderstand it." This statement is borne out by the reception to such translations at Oak Ridge, as reported by Zarechnak below.

The growing sophistication with regard to translation which is reflected in the book by Nida and Taber and in many recent publications calls for a new evaluation of the problem of machine translation, and a new statement on the current situation. The requirements for translation vary markedly from audience to audience. Even a glance at the Nida-Taber book, which concerns primarily human translations of the Bible, will disclose the difference between translation of religious and literary materials, and translation of scientific and technical materials.

For the translation of technical materials, the criteria of quality, speed, and cost have been used in evaluations. In the January Conference arranged under the Study, Bar-Hillel summarized his position on the improvements possible in machine translation in the foreseeable future using these three criteria. It is instructive to compare briefly these criteria with the objectives of Nida-Taber.

The primary concern of Nida-Taber is to "reproduce the message" of texts produced by cultures of the past for cultures of the present, often

radically different cultures, such as those of Africa and Asia. By contrast, the texts of interest to scientists and technicians share a common "culture," whether the texts are produced in Africa, Asia or in western countries. Zarechnak, who as director of Oak Ridge Russian-English translation is intimately acquainted with user needs of scientific translation, reports that American scientists readily understand translations of Russian scientific articles even when these are crudely rendered into English. Translations of technical materials accordingly face far fewer problems than do literary and religious translations.

This fact, obvious in any chapter of Nida-Taber, has often been disregarded when the requirements for technical and scientific translation have been discussed. Even articles and monographs dealing with machine translation have failed to be adequately explicit about the special problems of translating technical and scientific materials by computer. Instead, they have confused the problem by comparing machine translation with the long-practiced human translation, by equating the problems of translating scientific materials with those involved in translating literary materials, and by using the same evaluation criteria for the results. In his appended article Martin discusses the problem confronting the human translator of non-scientific material with respect to differences in cultural associations. The example he provides illustrates that some of these problems may not be encountered in technical and scientific materials.

Among the most striking contrasts are the disregard of cost and time in translating religious materials, as indicated briefly below.

The Appendix to Nida-Taber (1969, 174-188) deals with the "organization of translation programs." An organization consists preferably of three committees: (175) "1. the Editorial Committee, which has the basic responsibility for the work of translation,

2. the Review Committee, consisting of highly competent scholars whose advice and help is necessary, and

3. the Consultative Group."

After these three committees have made their contribution, a "stylist is called in" (1969, 186). This proposed organization, which is not untypical for academic projects designed to produce literary translations, provides perspective for the statements concerning post-editing of technical and scientific translations. Obviously, the length of time and the cost required to produce literary and religious translations are not factors of importance.

Yet timeliness is of increasing importance to users of scientific translations. Even in a relatively unhurried field like linguistics, few articles retain their importance over a long period. Statements have been made repeatedly about the obsolescence of publications issued a few years earlier. The insistence among technical specialists and scientists for speedy translation contrasts markedly with the length of time permitted for completing literary translations, and also with "the lag time (from receipt) in publication of the translated journals supported by NSF." This, according to a report of the National Academy of Sciences, (Languages and Machines, 1966, 17) "ranges from 15 to 26 weeks." This time span may be acceptable for archival purposes; for the requirements of scientists and technical specialists it may be burdensome.

Given a choice between overnight machine translation and human translation within two weeks, scientists at EURATOM invariably asked for machine translation. The need for virtually immediate translation is one of the major reasons for the concern with machine translation. In evaluating machine translation versus human translation, this reason may outweigh the difference in cost. And as Nida has pointed out, the parameter of "quality" varies considerably among the different users. Bar-Hillel, who some years ago coined the expression "High Quality Fully Automatic Machine Translation" now states in this appended article that he applied the expression in too absolute a sense. Further, that quality is related to the requirements of the user. This statement echoes the quotation from Nida-Taber on the shift of focus "from the form of the message to the response of the receptor." If technical experts and scientists have reasonable prospects of virtually immediate translation, the prospects may well be vigorously pursued, even if the translations will be more "labored" and "literal" than ordinary users permit for their religious and literary works.

In reviewing the prospects for machine translation, accordingly, the specific requirements must be considered as one of the major criteria. For technical specialists and scientists, translations must be consistent, reliable and timely, whether made by man or machine. Although the arrangements made for human translation are generally assumed to be known, and understood, a brief comparison of the current situation of human versus machine translation, and their prospects, may be useful before examining in detail the procedures involved in machine translation.

2. Translation: human and machine

The topic of machine translation is rarely discussed without reference to translation by man. In the comparison, several stereotypes have evolved. For clarity in dealing with the issue of machine translation these may be briefly noted.

The human translator is generally assumed to be highly skilled, both in the subject matter and in the source and target languages. Some commentators consider skill in the source language less essential than skill in the target language. Kay holds knowledge of the subject matter to be the most essential consideration. Accordingly, it will be no small task to provide machine translation systems with detailed information on scientific topics, and to program them to use this information. Human translators must also acquire knowledge of specific scientific and technical areas. With skills in the source and target languages, and control over the subject matter, the human translator is assumed to have great flexibility. Moreover, besides flexibility he provides immediate access to the text.

When, however, one considers the broad scope of scientific writing, and vocabulary, this ideal picture loses some of its attractiveness. The director of the translation service of the German government states flatly that no available dictionary is up-to-date. A translator dealing with German, one of the most thoroughly studied languages, would be unable to find any translation for thousands of technical terms. Others would have inadequate entries. Accordingly, technical terms might well be wrongly translated. A few moments of reflection by any specialist illustrates this problem of human translators.

One of the chief problems for any translator has to do with the changing meanings of supposedly standard technical terms. A simple area of linguistics, for example, is commonly subsumed under the terms: phonology, phonetics and phonemics. The term phonetics is generally used consistently. But among different authors the terms phonology and phonemics vary widely in their meanings. When such terms occur in texts translated by translation services, they are handled as though they had standard, fixed meanings. But for some writers phonology is used as equivalent to the term phonemics in other writers. Similar illustrations could be provided from any technical area, and all too easily supported by examples.

Without diminishing in any way the role of the human translator, we must conclude that adequate translation requires the organizational arrangements created to meet other contemporary technical problems. As noted above, an example of such an organization is proposed by Nida-Taber. If for a relatively unhurried problem in translation, teams of specialists are recommended, rather than individual translators, it is unrealistic to assume that an individual translator can deal with a broad scope of technical material. Moreover, even skilled human translators need retraining in expanding fields of science and technology if they are to keep up with new terms and new concepts.

To meet the problem, the German translation service has been compiling a large dictionary of technical terms and their standard translation. In this compilation, specific translations are fixed. The project, accordingly, is designed to standardize and normalize translations, as well as to provide assistance for human translators. Moreover, the dictionary is mechanized.

Eventually, any text to be translated is to be provided to the translator in a print-out having the translations of all terms in the dictionary, as well as the original. The translator's responsibility would then consist in framing the sentences in the target language. He would also determine the meanings of any new terms. In this way the dictionary would be expanded and updated.

The dictionary of the German translation service contains close to a million items. Problems which human translators face when using generally available dictionaries, which have far fewer entries, may be put in perspective by this resource. The arrangements for translators in the German translation service may also illuminate the requirements for computer-assisted translation.

It is occasionally proposed that computer-assisted translation is an attainable compromise, with better output than that from the individual translator and fewer awkward renditions than those provided by machine translation. Whatever one's reaction to this view, it should be noted that computer-assisted translation requires a large staff of research scholars, and a large computer facility. Kay, a proponent of machine-human translation, proposes an elaborate scheme to permit human beings to assist a system that is essentially a machine translation system. Under this scheme human beings would make decisions which the machine would be incapable of making and thus assure a high-quality output. His scheme envisions several native, possibly monolingual speakers of the source language, several monolingual speakers of the target language and one highly competent bilingual, to whom problems requiring knowledge of both languages would be shunted. In other words, the expenditure for staff

and equipment would not be small, actually larger than that for machine translation. Clearly, computer-assisted translation is proposed as a second choice, through desperation that machine translation is unattainable at present.

In contemplating machine translation, most observers have stereotypes which are as erroneous as are the stereotypes concerning human translators. The output of machine translation is supposed to be simply a printed document of some sort. Consideration of current computer technology however suggests a more likely output of a different kind. Many users of computers already have available display possibilities for research purposes, CRT's. If such research workers wish to secure a translation, little ingenuity would be required to provide it with great effect on the CRT, as described below.

The current arrangements at Oak Ridge, as described by Zarechnak in the January Conference, illustrate the potential output. Scientists at Oak Ridge who request a translation have the text keypunched and mechanically translated when computers are not fully utilized, typically at night. The translations are then available for them the following morning.

If, however, instead of a print-out, the translations were prepared for a running display on CRT's, both the original and the translation would be provided on the screen. The scientists would then be able to consult the original as well as see the translation. Tables and charts, often the most expensive sections to reproduce in translation, would need little attention under such a system. Zarechnak cited an instance at Oak Ridge where a

scientist detected an error by referring to a picture which accompanied the translation from the Russian. It turned out that the Russian original contained the error which was carried over into English. The availability of pictures, tables and charts thus provides a check on accuracy. Other advantages of displays of translations on CRT's will not be pursued further here. One obvious advantage is the speed with which the translations would be provided.

Among the advantages of machine translation is consistency. As in the German translation service, standard terms could always be produced. As a simple example, the German translation service decided to use Telefon rather than Fernsprecher; even the variant Telephon was considered erroneous. In much the same way, any technical term need never be varied, unlike the practice of many translators.

If the quality of such translations is to equal that of the most accurate human translations, a comprehensive dictionary and grammar are essential, as well as the necessary hardware and the software techniques. Achieving these has been the major goal of machine translation. In the next section we note the current status of these three requirements.

3. Techniques involved in machine translation: hardware, software, linguistics

In the early attempts at machine translation, the capacities of computers were a major problem. Difficulties resulted especially from the inadequacy of rapid-access memories. For processing languages, the available rapid-access storage space was filled with the major rules for grammatical constructions. Lexical items accordingly had to be stored in memories, usually on tapes, which required a considerable period of searching. As a result, even simple sentences required a long time for analysis. When the Linguistics Research Center was carrying out its research with an IBM 7040, several years ago, the computer would "grind" all night to translate a few sentences.

To speed up the process, attempts were made to develop special-purpose computers, used only for machine translation. A notable example was the Mark II. Intended for only a short period of use, it was long maintained because of the inadequacies of general purpose computers for data processing, including the procedures required for machine translation.

Remarkable improvements, especially in rapid-access storage devices, have largely eliminated the problems caused by inadequate computers. Lexical items can now be retrieved as rapidly as were the major syntactic rules a decade ago. And with further improvements of storage devices in process, computers no longer pose major problems in machine translation. Proper management can assure the necessary space for storing large sets of lexical and syntactic rules. Skillful use of them depends in great part on recent

software developments.

Programming has evolved as rapidly as have computers. The early higher level programming languages were designed largely for application to numerical analysis oriented problems. Linguists who used computers were compelled to write their own assembly language programs. Eventually, in response to the needs of computational linguists, procedural languages such as COMIT were developed; in addition, other higher-level symbol manipulation languages such as LISP and SNOBOL proved useful for linguistic applications in mind. A key factor here was the enrichment of programming language data types which made possible efficient representation and manipulation of linguistic structures.

In time, procedure-oriented languages were used to produce programs of more general usefulness to linguists. Dictionary lookup and maintenance programs and context-free grammar parsing programs were followed by such programming systems as J. Friedman's transformational grammar tester and S. Petrick's transformational grammar syntactic analyzer. Systems such as these can be considered to be problem-oriented programming languages. The IBM natural language question answering project mentioned by Petrick in his appended paper uses Friedman's grammar tester system as well as a transformational syntactic analysis system that provides for a linguistically more realistic class of transformational grammars than could previously be accepted using its predecessor.

Certainly presently existing procedural and problem-oriented languages make the mechanization of many linguistic processes easier than was the case

a few years ago. The programming of many linguistic algorithms remains a slow and difficult task, however, as is the case for most complex algorithmic processes.

The scope of the programs necessary for machine translation may be noted by examining a flow-chart of the programs that had been projected, and in part completed, at the Linguistics Research Center. These were produced entirely from scratch. Because the basic programs furnished by computer manufacturers were so inadequate, the Linguistics Research Center programs were written in machine language. The expenditure of time was enormous. The magnitude of the problem may be noted if one compares the cost incurred by IBM in developing PL-1; the cost of it far surpasses the entire amount which was spent on machine translation from the beginning of machine translation research.

Gradually, adequate computer programs were devised for data processing. These now form the basis of programming systems used for machine translation. Like all programs, they need modification, and improvement, especially to speed up processing. The basic programs, however, are available and to them the additional programs needed for language processing can be added. Like computer equipment, programming systems will be improved. But in the same way, they no longer provide an obstacle to work in machine translation. Kay, an authority in the field of machine translation software, states flatly that "the real problems of machine translation are not in program writing; they are in linguistics."

All participants in the study agreed that the essential remaining problem

is language. Language has proved to be immensely complex, far more complex than linguists themselves had recognized. Moreover, as noted below, linguists do not agree on what constitutes complexity in language.

The fundamental linguistic problem for machine translation is often discussed in terms of deep or underlying structure. Numerous examples have been given to illustrate the problem; among the most commonly cited are the sentences:

1. He is easy to please
2. He is eager to please

Both sentences are alike on the surface. Yet their meanings differ, as the paraphrases indicate:

1. Someone pleases him readily
2. He pleases someone else with alacrity

Speakers of English interpret each sentence correctly. A machine using an inadequate model of English would not. It would take "he" to be the subject of the verb + adjective combination, and also of the infinitive; its analysis of sentence 1 would therefore be wrong. A literal interpretation would fail to determine the proper meaning of this sentence and many other sentences. By determining the deep structure the meaning can be more easily arrived at. Alternatively, "easy" may be provided with a feature which would transform "NP be easy to -Inf" to "it be easy to -Inf NP" in accordance with Harris' use of transformations.

Language is structured in this way in all its components, the phonological as well as the syntactic. In both of these components, it is a code,

rather than a cipher, system. The human brain knows how to interpret the code. If machines are to interpret language, they must be provided with a comparable capability. Engineers have been working on machine interpretation of the phonological system; it would be useful, for example, if telephone "dialing" could be done by voice, rather than manually. Engineers have not mastered the problem, however, even though they are aware of the basic difficulty.

The problem in the syntactic component of language has been one of the central issues for linguistics since the publication of Saussure's Cours in 1916 (though it was known earlier). Various labels have been given to the underlying structure. Saussure used the traditional philosophical terms: "form" for the underlying structure and "substance" for the surface structure. Recently the term surface structure has been used almost exclusively rather than substance, and deep or underlying structure rather than form.

In view of this structure of language, techniques must be devised to get from the surface structure to the deep structure. Recently linguists have proposed to do so by specifying the relationships between surface and underlying structure by means of transformations. Current linguistic descriptions state these relationships in such a way that the abstract underlying structures are transformed into the surface structures. That is to say, descriptive grammars start from deep structure and relate it to surface structure. If, however, transformations are applied in computer analysis of language, the deep structures must be primarily determined from the surface structures. Accordingly, for the computer, reverse transformations must be devised.

These would not simply be the obverse of transformations linking deep structure to surface structure.

The technique of using reverse transformations was explored almost a decade ago. But it was unsuccessful, primarily because the proposed transformations yielded too many underlying structures, including wrong ones for any given surface structure. In devising reverse transformations, the linguists had not been able to refer to lexical features. The system being developed at the Linguistics Research Center can make use of lexical features, and thus is meeting this problem. Descriptive linguists have also encountered the problem of producing too many alternatives by means of transformations, and are painfully aware of it. Transformations are now generally acknowledged to be too powerful (Bach; Peters and Ritchie, 1971).

This realization has important consequences for machine translation. The fact cannot be escaped that in machine translation, one must somehow determine the underlying forms of sentences. Further, the technique of using some relational formulae like reverse transformations is also clearly necessary. In considering linguistic techniques, the fundamental question is: how can these formulae be adequately restricted so that they yield only the specific underlying structure intended by the author, that is, the proper meaning?

Two devices must be used: the lexical elements must be described as precisely as possible, so that only the desired transformations apply; the transformations must be devised in such a way that their use is properly restricted.

Exploiting this understanding of the necessary procedures will require considerable work. The lexical analysis alone will be a huge task. It will,

however, lead to vastly improved general purpose dictionaries, and vastly improved understanding of language, with implications for various applications. Jakobson confirms the inadequate state of dictionaries in all languages. Developing the necessary techniques to arrive at underlying structures will also require considerable study.

Like programming techniques, and computers, the linguistic techniques will continue to be improved. Unlike them, the techniques are not yet available for any appreciable amount of any language. As stated above, dictionaries are inadequate. And the fullest account of the grammar of any language: Integration of Transformational Theories on English Syntax, by Robert P. Stockwell, Paul Schachter and Barbara Hall Partee, is generally considered premature. For these reasons, specialists state that linguistic analysis is now the major problem if machine translation is to be achieved. The state of linguistics is accordingly of vital importance for machine translation.

4. Contributions of linguistics to machine translation

Two important topics of recent linguistic research have been of major concern for machine translation, and have been in part prompted by work in machine translation: the problem of a universal base, and that of ambiguity in language. Views on each of these topics have changed considerably, even during the last decade. Linguists also differ in their views on each, as the following statement may indicate.

The assumption of a universal base receives support from the capability of speakers to translate. It is also supported by the capability of infants to learn any language, to learn it rapidly, and in accordance with well-determined stages. Whatever a baby's ancestry, it acquires the language it hears. Moreover, the stages of linguistic development are fixed for virtually all infants, regardless of their intelligence.

These observations are most plausibly accounted for if we assume some fundamental principles common to all language; further, that these somehow are related to the functioning of the brain. The principles are highly abstract. They permit certain linguistic structures and constrain others which are theoretically possible. As yet they are not by any means thoroughly explored. The term "universal" has been used for general characteristics of language; one example of a universal may be exemplified here, with two sentences and their variants.

1.a She regretted the fact that she had taken the book.

1.b She regretted the fact that she had taken what?

2.a She regretted that she had taken the book.

2.b She regretted that she had taken what?

2.c What did she regret that she had taken?

BUT NOT 1.c *What did she regret the fact that she had taken?

The impossibility of 1.c results apparently from a universal principle which blocks the extraction of an element out of a clause modifying a noun phrase. This principle was formulated by Ross (1967, 66-70) as "The Complex NP Constraint." Since this principle applies to all languages which have been examined, it is assumed to be a universal characteristic of language.

Whatever the views which will be formulated concerning universals, this principle, like other universal principles that are being investigated, restricts the possible transformations for structures of language. Since language is governed by such constraints, the model which must be constructed to embrace all languages must have certain limits. Moreover, if only because of the finiteness of the human brain, grammars must be finite.

These observations lead to the conclusion that a mechanical translation system can be devised. Even more support is provided by the conclusion of much recent linguistic study that we may posit the existence of a universal base. For the surface structures of any language can be related to such a universal base. Since the universal base in turn can be used for deriving the surface structures of any language, the universal base can serve as the intermediate language between any source language and any target language.

The possibility of devising a translation system in view of the fact that a universal base may exist still leaves many problems. The surface structures of one language may map into the intermediate language differently from those

of the target language. For semantic distinctions which are overt in one language may not be overtly expressed in another language. Different languages may also represent differing "word views." We assume, however, that closely related languages, like English and German, are similar in expressing their semantic distinctions overtly and covertly, and even in their surface structures; accordingly, they are relatively easy to translate into each other. Moreover, languages strongly influenced by another language, as were the languages of western Europe by Latin, so that Whorf referred to them as Standard Average European, share many surface features. Accordingly, translation systems may be so designed that the description of any source language is directed at a specific target language. And in our current understanding of language, attempts to move directly from any source language to a universal base, in the hope of translating into any other language, are premature. For we must take into account the complex relationships between a given language and the universal base.

That linguists have not yet been able to determine the exact nature of a universal base does not present a problem for linguists who favor machine translation by means of an intermediary language. Garvin has stated (1970, 9-11) that an intermediary language need be nothing more than a series of symbolic notations to record the output of the recognition routine for the source language and to serve as input into the command routine for the target language.

Further, as mentioned in the preceding chapter, rigorous procedures for establishing relationships between surface structures and a universal base must be established. For example, as Ritchie and Peters (to appear in

Information Sciences) have demonstrated, the transformations characteristic of most generative grammars are so powerful that they permit the derivation of any structure. Consequently, in spite of their contributions to our understanding of language, they can make no claim on properties of grammars, on a universal base, or on what is going on in the brain. Current transformational grammars are accordingly inadequate devices for describing languages and also for use in machine translation systems, for which they produce far too many syntactic interpretations of any given sentence.

The production of devices to map surface structures stringently into underlying structures is one of the most serious concerns of current linguistics. Bach's paper noted above (1971) is an example. A device projected by Stachowitz has been described in RGEMT (1970). It makes use of an underlying form, the standard strings of a language. Associated with these strings are canonical forms, which represent the meanings of given sentences. "The language" of these is assumed to be "common to all natural languages" (Stachowitz, 1970, T-65). Fuller information on the model is given T-66ff. The quotation here may be adequate to indicate that the canonical forms correspond to a universal base. A description of translation as it is being pursued in accordance with this model at the Linguistics Research Center is appended (Stachowitz paper).

Of great importance for the Linguistics Research Center system is a well-designed lexicon. The intensive lexicographical work which has been going on at the Center for more than two years now has resulted in great amounts of syntactic information; the incorporation of semantic information is currently in progress. Because of their syntactic and semantic classification, the lexical

entries will limit the possibilities of relationship with canonical forms. In this way a proper match will be brought about between the lexical and syntactic elements of the source language and those of the target language.

The design of the lexicon has been vastly improved over dictionaries envisaged a decade ago. Without intending to dwell on the naivete of these and their proponents, reference might be made one further time to the saying which was supposedly quite ambiguous and accordingly a prime exhibit of the difficulties of machine translation: "Time flies like an arrow." When one notes that Austin's How to Do Things with Words was available at the time this saying was widely discussed, the singular limitations of a view of linguistics which could permit the citing of the proverb seem quite remarkable. For as Austin made clear, meanings cannot be determined from syntax alone. Additional publications, such as Speech Acts by John R. Searle (1969), make the simplistic attention of a decade ago to a highly limited type of syntax quite difficult to understand.

As linguists have improved their models of language, the problem of ambiguity has been reduced. It may be noted that the attention to the pragmatics of Peirce repeats a position held early in machine translation research. By attention to pragmatics, that is, to information on the "origin, uses and effects" of language, sentences belonging to the class of proverbs are not treated like sentences found in scientific exposition. Attention has also been focussed on such classes as illocutionary verbs, or on the characteristics of speech acts. Thus a sentence like: I pronounce you man and wife would not be treated as a simple declarative statement, with the meaning of "pronounce" in a sentence

like: They pronounce greasy with a voiced groove fricative.

While models of language in this way incorporate far more information about individual sentences than did the purely syntactic-based grammars of a decade ago, means must be devised to take account of the more accurate analysis of language which is now projected. Suggestions vary concerning the implications of these developments, as the following chapters will indicate. Some specialists consider machine translation unlikely unless at the same time automatic information and fact retrieval are made possible. Others hold that machine translation is not now, and may never be, contemplated for types of language outside technical and scientific documents; accordingly the "origin, uses and effects" of the material to be translated are determined, and investigators dealing with machine translation should direct their concerns at this restricted type of language.

Whatever steps are selected to employ findings of contemporary linguistic to carry out machine translation, it should be noted that specialists in machine translation have taken account of these findings. There accordingly is no conflict between specialists in descriptive linguistics, linguistic theory and machine translation, as Chapter 6 below will outline in further detail. As descriptive linguists improve their understanding of language, and the models by which to express that understanding, machine translation specialists will update their procedures and models.

5. Pertinent recent work in linguistics

Contemporary linguistics is concerned with all facets of language, its syntax, semantics, aberrant uses, uses in established social situations, its relation to other disciplines, such as logic and so on. This breadth of concern contrasts strikingly with self-imposed limitations of the recent past. It also leads one to examine the extent of concern of machine translation with these matters. For machine translation is an application of linguistics, like language teaching, design of communication channels, and the like. Each of these applications relies in part on disciplines other than linguistics. Each may or may not be involved with the various sub-disciplines of linguistics.

To what extent applications are intricately tied to the various concerns of linguistics is a highly important problem; for many of the most active theoreticians in linguistics deplore their inadequacy in understanding language. Some state that an understanding is not foreseeable for a considerable length of time, such as half a millennium. Moreover, there is no comprehensive linguistics theory. Most of the current efforts of linguists, even those who style themselves theoreticians, are directed at minutiae. These efforts, such as the repeated examination of quantifiers like <u>some</u> and <u>any</u>, are carried out chiefly to find support for new hypotheses. If adequate language teaching, or machine translation, or any other application of linguistics is so dependent on a thorough understanding of language that it must be deferred until linguists arrive at a satisfactory linguistic theory, anyone with an interest in these fields should be advised of his dubious expenditure of time, energy and funds.

In dealing with this important question, we may examine the concerns

of those linguists who are especially vocal about our lack of understanding of language. John Ross furnished an excellent example of these concerns in one of his presentations sponsored under the Study. It will be summarized briefly here to illustrate topics which engage the interest of linguists, and their possible pertinence for machine translation.

For some time linguists have been concerned with the role of quantifiers in language, such as some : any. Among such recent linguists are Brugmann, Jespersen, and Sapir. Their discussions have not solved all of the problems with quantifiers. Current concern extends beyond the simple uses, such as those included in Fowler (1965, 31): "Have you any bananas? No we haven't any bananas. But yes we have some bananas." This quotation implies that some is used in positive contexts, any in negative, even though it is provided in Fowler's discussion of any (= some) in positive sentences like: "Then, for the first time, she paid any attention to my existence." While Fowler's entries for any and some concentrate on stylistic matters, contemporary linguists deal with the syntactic and logical problems. For example, they may note that in addition to other restrictions that cannot be included here any is not used before verbs with negative implication, such as doubt, dislike, and predicate adjectives such as unkind. For example, the following sentence is impossible: * Anyone doubts that the earth is flat. But the following sentence is possible: Anyone would doubt that the earth is flat. Similarly, in contrast with the impossible: * Anyone is unkind the following sentence is possible: Anyone who fails to send his mother a card on Mother's Day is unkind. The precise constraints have not yet been defined.

One of the problems in this definition is the existence of two meanings of any, one illustrated in Ross's sentence:

Anybody could have shot Max

The meaning of any here might be made more precise by adding: whatsoever. This use of any is found only when possibility is involved; it is not used with must. A different meaning (some) is found in the question: Do you know any songs? If, however, stress were put on any, especially in a negative sentence, the "any. . .whatsoever" meaning would emerge.

Accordingly, as study of quantifiers has been pursued more extensively it is clear that a sentence like the following can have two meanings:

We do not believe that any catalyst could have precipitated the reaction.

On one interpretation, this sentence could be roughly equivalent to: that some catalyst, that is, that a selected catalyst was involved. By another interpretation it would be equivalent to: that any catalyst whatsoever was involved, that is, that no involvement by a particular catalyst was possible.

This indeterminacy of usage in English presents a translation problem, for German, too, indicates slight differences of meaning with quantifiers such as irgendein 'someone, anyone'. These differences are especially important in the colloquial, as the following quotation from the Duden Grammatik (1959, 265) may indicate:

Irgendeiner muß es doch getan haben!

"Someone must surely have done it!"

Such usages are often found in conjunction with modals and adverbs, like doch in this sentence. For the examples of any which Ross cited, however, translation

would not be a problem inasmuch as the ambiguity in English is preserved when literal translation into German is carried out.

The use of any in such a relatively straightforward sentence is simply the beginning of Ross's interest. He has pursued the difference in uses of any in syntactic constructions illustrating general syntactic patterns or principles which he has investigated intensively. One of these is a "maximal domain of syntactic processes, "—in his word "island".

An island is subject to special constraints, as the impossibility of the following example indicates:

*Who do you believe their claim that Max fired?

If "their claim" were eliminated, the equivalent sentence is possible:

Who do you believe that Max fired?

By the use of claim, an island is created which does not permit the extraction of the object of fired. See Chapter 4 above.

In his discussion, Ross proposed that islands have an effect on the meaning of any, e.g.

We do not believe their claim that any catalyst could have precipitated the reaction

For Ross, any in such a context can only mean "any. . .whatsoever". Not all participants in the Conference agreed. Yet the situation is presented here nonetheless to illustrate the delicacy of concern which occupies current syntacticians. Readers may decide for themselves the importance of possible ambiguities in such sentences if machine translation were confronted with them.

Another syntactic principle explored by Ross is in his words the domino effect. By this principle, selected elements like deny affect subsequent elements in the sentence. Thus, deny rules out the last of the following sentences even though the others are possible:

1. I deny that anybody said anything nasty
2. I deny that anybody said something nasty
3. I deny that somebody said something nasty
4. *I deny that somebody said anything nasty

In the same way, according to Ross, any cannot be used in lower clauses when some is used in upper clauses, as in:

Finding somebody under the bed is not easy for some people

That is, if the first phrase read: finding anybody under the bed, the sentence would be impossible, on Ross's view. On the basis of such patterns, Ross asserts that domino constructions disambiguate. Accordingly, the following sentence can only have the "any.. .whatsoever" interpretation:

We did not inform some researchers that any catalyst could have precipitated the reaction

From these examples, Ross concludes that in machine translation the sense of any would not be determined until the machine had searched for some in the same sentence. That is, machine translation would be faced with a large job. For the meaning of individual words could not be determined simply by examining the immediate contexts. Rather, entire passages would have to be examined. The problem would be compounded because linguists have not yet determined the extent to which individual areas contribute to disambiguating meaning; see Chapter 6.

Ross cites these and other problems in support of his contention that linguists have a highly inadequate view of language. The extent of his pessimism may be illustrated by his statement that the sentence: Birds sing is full of a host of unsolved problems.

As noted above, individuals will have to evaluate our current understanding of language and determine for themselves whether Ross's bleak characterization of the state of linguistics implies that machine translation now is impossible. The extent to which machine translation depends on the production of a comprehensive linguistic theory was discussed repeatedly during the Study. Obviously the sentence: Birds sing can be readily translated: Vögel singen, whether or not Ross's problems are solved. In much the same way, many complex constructions may be translated straightforwardly, especially between closely related languages.

It may not be presumptuous then to suggest on the basis of Ross's problem-riddled example that man and machine can happily embark on translation projects, even before current and future generations of linguists solve its momentous problems.

For students of translation, a second paper may furnish more comfort concerning the contributions of linguistics—the paper of Fillmore which undertook to determine the possibility of finding an unambiguous interpretation for the sentence: May we come in? Fillmore selected the sentence, partly because it is extremely barren in providing assistance to a hopeful interpreter. It contains no content word for the subject, as does the sentence: Birds sing, to delimit the meaning of actor. For like other auxiliaries, may has a broad

scope of meaning. Yet Fillmore demonstrated that proper use of all linguistics features leaves little question about the proper meaning of the sentence.

Since the paper is appended, it may be consulted for details. The general types of analytic techniques, as well as the types of "linguistic information" in this sentence pointed out by Fillmore will be summarized here. Moreover, in evaluating the implications of Fillmore's analysis for machine translation, it is important to note that he used only the written form of the sentence, excluding information that might be obtained from "any understanding of the voice quality of the speaker on the manner of utterance."

Identifying first the "syntactic information" in the sentence, Fillmore uses it to determine among the three possible functions of may the one which is appropriate in this sentence.

Next, examining the "illocutionary force of the question," Fillmore notes the information on deixis furnished by the pronoun and the verb come. The term "illocutionary force" refers in Fillmore's paper to the obligation which the question imposes on the addressee, that is, the obligation to exercise authority. The term "deixis" refers to the various aspects of the interpretation of sentences that relate to the speech act situation, such as person deixis, place deixis and time deixis. The possible meanings of come are restricted by its use in "a permission-seeking utterance."

Last, it may be noted that Fillmore determines the meaning of the sentence from its "surface structure." He has done so by using a comprehensive lexical description for each of the four words. The possible meanings of each

are restricted by the order of the sentence and by the selection of the other elements. That is to say, disambiguation was carried out by using two syntactic devices: order and selection.

In conclusion, Fillmore lists "the various kinds of facts which must.. . be included in a fully developed system of linguistic description." These are extensive. Yet such explicit linguistic descriptions permit a mechanical disambiguation, and interpretation, of a given sentence.

The effort required to produce these descriptions will, however, be enormous. An example of the analysis necessary for improved interpretation of sentences, which will be particularly important for information processing, is Karttunen's paper, "The Logic of English Predicate Complement Constructions." This paper, which is also appended, leads to seven classes of verbs, each indicating a commitment which "the main sentence carries along.. .with respect to the truth or falsity of its complement" and an indication of "what is implied." For example, the verb cause belongs to one of these classes which carries a commitment "true" for main as well as complement sentences. The seven classes of verbs arrived at in the paper identify meanings in much the same way as did the syntactic information in the sentence: May we come in?

Linguists accordingly are drawing nearer to lexicographical work of the past, as represented especially by Zgusta and Josselson in the Study. Since the use of lexicographical techniques for machine translation is discussed in the appended papers of Zgusta, they will not be further noted here.

Current linguistic description in this way is providing information on detailed lexical classes, as well as on syntactic constructions. These two types

of information about language, whether they be labeled syntactic or semantic, are leading to descriptions of language which are so precise that the sense of a sentence can be determined mechanically.

The generative semanticists, besides Ross, who participated in the Study: George Lakoff, Robin Lakoff and James McCawley, are also contributing insights which will sharpen our understanding of language. McCawley, for example, challenged the analysis of adjectives as reduced forms of relative clauses. In his presentation he cited examples, such as: He's an incredible fool which cannot be derived from: He's a fool who is incredible. Further, while the sentence John is an easy man to please is acceptable, the construction is inacceptable for a noun like that in the sentence: *John is an easy tool-and-die-maker to please. Such observations point out the necessity of distinguishing sub-classes of "adjectives" and carefully defining their uses in somewhat the same way as Karttunen and other linguists are distinguishing sub-classes of verbs.

These linguistic analyses will not however solve the problem of sentences which Austin labeled "performatives". One of the participants in the Study, Fraser, concerned himself particularly with the problems raised by Austin, and pursued by other philosophers such as Searle. As Austin pointed out, performative sentences contain "humdrum verbs in the first person singular present indicative active." An example is: "I name this ship the Queen Elizabeth." Few linguists are unaware of Austin's book How to Do Things with Words (1962). It is noteworthy that the book concludes with a sub-classification of the verbs discussed. There are five classes:

1) verdictives, 2) exercitives, 3) commissives, 4) behabitives and 5) expositives. The example given above belongs to Austin's class 2. Like the work of the generative semanticists, Austin's leads to more precise lexical analysis.

Modestly, Austin does not claim to have produced a definitive classification, but includes in his final chapter the statement: (1962/68, 148) "Now we said that there was one further thing obviously requiring to be done, which is a matter of prolonged fieldwork." Generally, fieldwork is used by linguists of work in non-literate cultures. Austin obviously recommends highly exacting analysis of the types discussed earlier in this chapter.

After Fraser's presentation, the problems raised for machine translation by illocutionary verbs were discussed, especially by Ross and Bar-Hillel. Like Fraser, Ross stated that a theory of speech acts is not essential for machine translation, inasmuch as its goal is the translation of technical materials. Agreeing with this view, Bar-Hillel added that nonetheless we need a great deal of pragmatic information.

This chapter was intended as a survey rather than as an evaluation of current linguistic work as pertaining to machine translation. Linguists are carrying on research into the narrower concerns of language, and those listed in the first paragraph of this chapter. Though not all of this work may be directly pertinent for machine translation, specialists in machine translation are profiting greatly from it. The implicit interest of this work—to analyze sentences so thoroughly that they can be interpreted from the linguistic information contained within them—is of great significance for machine translation.

To the extent that this interest is accomplished, machines wilt be able to translate.

6. Views of specialists concerning machine translation

One of the primary problems in presenting the views of specialists in machine translation results from the low level of research during the past five years. Few groups received any kind of support. The greater part of them could only update their previous systems, not introduce major innovations. In view of the low funding, research was severely restricted, generally devoted to improvements in the lexicon. This limitation in funding greatly restricted the possibility of carrying out new experiments, let alone that of producing improved translation systems which could meet some of the goals held out for machine translation. The views of specialists are accordingly based in part on assumptions framed some years ago when some long-range machine translation projects were able to carry out work in programming and in linguistic analysis, and to test their efforts by means of computer runs.

In his summary on the final day of the January Conference, Bar-Hillel concentrated on the linguistic situation. Noting that the primary considerations are quality, speed and cost, he expected improvements in speed and cost of output from advances in computer hardware and software; but their contributions to improved quality would only be external, for example as printouts would begin to approximate those produced by printing-presses. Essential for improving quality is improvement in linguistic theory and analysis.

Bar-Hillel's discussion involved arguments on a definition of quality, and on the receptivity of scientists to output from the translation systems which now are in use, notably the Georgetown system as used at Oak Ridge. This point will be discussed further below, in connection with Zarechnak's statements.

Bar-Hillel disagreed with Zarechnak's statement that the output was approximately 80% complete. His own estimate is around 35 to 40%. He assumes that sustained linguistic work during the next five years can raise this figure by 15%. And he considers this degree of accuracy less than that which sponsors demand. But as a perceptive participant at the Conference pointed out, Bar-Hillel's estimate has to do with readability. Zarechnak's on the other hand has to do with informativeness, as the reaction of the Oak Ridge scientists indicates.

Turning from this pessimistic estimate to possible palliatives, Bar-Hillel rejected the possibility of simplification of scientific texts by editors, and he also doubted that scientific texts presented many fewer problems than do general materials. Among the questions discussed during Bar-Hillel's summary were inadequate means devised to measure complexity. Fraser pointed out the difficulties involved in determining such means, for complexities may exist at the various levels of language, by no means in parallel ways. The question accordingly is another which has been raised in machine translation research.

In what he considers the present impasse, Bar-Hillel proposed that fully automatic machine translation must be sacrificed for the time being, and that instead efforts should be made to develop man-machine combinations for translation. He had no suggestion on the type of combination, but indicated that the most promising one would have to be determined by research. During the discussion Pankowicz pointed out that no translation, indeed no writing, is published without editorial intervention. Bar-Hillel envisages more "human

intervention" than is the normal practice in translation. He concluded by suggesting that determining a useful type of man-machine translation system would not be difficult.

However negative he is to fully automatic machine translation at present, Bar-Hillel recommends that further research be carried out. In his view, such work is intellectually respectable and challenging. It is of further interest to him because of its potential contributions to information retrieval.

Garvin is far less pessimistic about the prospects for machine translation. In his view, the basic problems involving linguistic analysis have been identified. What remains to be done is application, involving detailed lexical analysis. By Garvin's view, if adequate funding were provided, acceptable outputs of machine translation could be achieved in five to ten years.

Neither Garvin, nor any of the other participants, however, were able to define what is meant by acceptable translation. Moreover, Garvin believes that linguists must leave such determination to users. A similar position was maintained by Kay, who stated that "buyers" of translation are the persons best equipped to pass judgment on the acceptability of translations. A comprehensive judgment accordingly is unlikely until machine translation is carried out.

Garvin's views on the essential problems are indicated in his appended paper. For him too, the linguistic problem is primary. Moreover, a system must be prepared in accordance with an adequate model. He states further that all systems under development are of a tripartite design, but that grammar and computer programs may be linked. Specific problems, in his opinion, must be solved in terms of the overall system. There can be no hard and fast

rules, for example, for deciding whether to handle any particular topic in the grammar code or in the algorithmic portion of the system. Garvin is also flexible in his recommendations for performing analysis; he does not advocate rigid analysis from left to right, but rather selective analysis. As an experienced practitioner in the field, accordingly, Garvin believes in making use of any advantages which can be offered by linguist or programmer. The designer of a translation system relies on them for any possible assistance, but like any applied scientist he must decide when their contributions are useful in helping attain his ends.

Of particular interest is the IBM question answering system mentioned in Petrick's appended paper. This system is based on a generative transformational syntactic component that reflects current transformational theory to a reasonable extent. Deep structures are in some cases quite deep; certain nouns for example, are transformationally derived from underlying abstract verbs. The use of relatively deep structure facilitates semantic interpretation; this is accomplished through the use of a translation mechanism due to Knuth. (The task of relating these deep structures to surface forms is, to be sure, quite complex. Even relatively simple sentences may require as many as forty or fifty transformational applications.)

The syntactic analysis algorithm which is utilized is valid for a significant class of transformational grammars. This, together with the modular nature of the Knuth semantic interpreter, makes modification of both the syntactic and semantic components relatively easy.

It should be noted that the system being implemented at the IBM

Research Center is an experimental question answering system based on a restricted subset of natural English, not a machine translation system involving relatively unrestricted textual material in two or more natural languages. This choice of application was made because the Theoretical and Computational Linguistics Group at the IBM Research Center feels that the coverage of English presently attainable by any means is too small for the purposes of machine translation but perhaps not too small for applications which can tolerate artificially restricted input. It should also be noted that the IBM system is in an early state of development with the separate components (generative transformational grammar, syntactic analyzer, and semantic interpreter) yet to be integrated and put to use as a coherent whole.

By contrast, the system used for machine translation of Russian at Oak Ridge is essentially the GAT system developed at Georgetown University some years ago. The lexicon and syntactic rules are updated by Zarechnak constantly. But the basic system, based on a surface structure analysis, has been maintained. As indicated above, the system is frequently used. According to report, the output is valuable, though scientists must become used to its Russian-like syntax. If translation indicates that a paper is particularly valuable, a further effort can be made to produce a completely accurate version of the original. Zarechnak considers the output of great value to the Oak Ridge scientists. They must also find it useful, for numerous translations are produced on the initiative of individual scientists.

As several participants in the Study pointed out, a convincing analysis of the usefulness of machine translation will result only from a well-designed

experiment determining reactions to its output. Yet linguists have identified the stages of accuracy, in accordance with the levels of language which must be incorporated in a system, as shown in the examples below. It is unlikely that scientific or technical articles would contain such simple sentences. Yet on the assumption that no further linguistic information were available in the sentences to be translated, these examples illustrate successive degrees of sophistication in the development of machine translation and the expected quality at each stage.

Proceeding from the simplest system to one which approximates the information available to a human translator, we may propose the following progression:

translation with:
1. Lexical information alone
2. Syntactic information
3. Semantic information
4. Contextual information
5. Pragmatic information

Each state incorporates all of the earlier stages.

The following examples indicate difficulties, and characterize inadequacies, which each type of system fails to resolve. If that type of system were used, these shortcomings would have to be removed by pre-editing or post-editing.

1. Lexical translation, with no access to syntactic information.

They milk cows.

Under such a system milk might be taken as verb or noun.

2. Syntactic translation, with no access to semantic information.

The conductor broke.

The conductor smiled.

Under such a system disambiguation would be impossible.

3. Semantic translation, without contextual theory.

We watched the conductor. He smiled.

We watched the conductor. It was on fire.

Here too disambiguation would be impossible.

4. A system making use of contextual information would be able to disambiguate the sentences given in 3, and other examples, such as the German verb beugen in the indicated contexts.

beugen = bend

= deflect (optics)

= inflect or conjugate (grammar)

5. A system with access to pragmatic information would provide for the German sentence "Eisenhower folgte Truman" the correct reading "Eisenhower succeeded Truman" rather than the equally correct alternate translation "Eisenhower obeyed Truman."

In accordance with this sketch of potential systems, we expect the highest quality from a system which is at stage 5, or possibly at stage 4. The requirements for these stages have not yet been handled in linguistic theory, and accordingly at present they are unattainable. To what extent a system at stage 3 will be able to translate scientific and technical materials acceptably will depend on testing of the output, and the receptivity of users after such a system

has been developed. Systems at this stage are now under development.

The questions raised in the Study are also of interest to scholars who could not participate, as a recent article by Kulagina, Mel'chuk and Rozentsveyg indicates. It is noteworthy that, like Bar-Hillel and other participants in the Study, the three authors concentrate on the quality to be achieved, assuming that cost and time can be adequately managed.

The authors express their views concerning the feasibility of machine translation with regard to the ALPAC report, especially its view that machine translation is at present impractical. They state: "We wish to declare decisively that this view has no real support: it is founded upon a failure to understand the problem in principle and confusion of its theoretical, scientific and practical aspects. The fact that machine translation has been ineffectual in practice to the present should, in our opinion, lead to an increase rather than a decrease in efforts in this area, especially in exploratory and experimental work. It is clear that no practical result can precede fundamental development of the problem, although the possibility is not excluded that useful practical results may be the product of early stages of research. There is not, and has not been, a crisis in machine translation as a scientific undertaking, a crisis which would be reflected in a lack of ideas and a lack of understanding what path to follow. Machine translation as a scientific undertaking.. .is continuing to develop actively. There are many interesting ideas and approaches which are far from being sufficiently developed and experimentally tested."

After making this critique of a negative approach, they state that a high-capacity, high-quality system can be established within about five years

if (a) the components produced under a system—dictionaries, grammars, algorithms—are prepared on the basis of available theoretical developments and experimental results, (b) goal-directed experiments are carried out (given an existing algorithm), and (c) research oriented toward a future modification of the system is carried out.

Having arrived at such a judgment, the authors suggest the establishment of an MT center comprised of a linguistic, a mathematical and a computer group equipped with the best computational equipment available. From such a center they expect not only contributions to machine translation, but also to automatic language processing in general. They also believe that if their recommendations are followed, machine translation on an operational scale can indeed be accomplished within their estimated deadline (4 five years).

The estimates, and recommendations, of this article coincide remarkably with those of some of the participants in the Study, notably Garvin. Readers may form their own judgments on the basis of the appended articles, not all of which agree in these estimates or recommendations. Like Garvin, Kulagina, Mel'chuk and Rozentsveyg have been involved in the theoretical and practical aspects of machine translation. Their views are accordingly based on intimate knowledge of the difficulties involved in attempting to achieve high quality machine translation.

Conclusions

1. A technological application drawing on three bases: computer hardware, computer software and linguistic analysis, machine translation today is confronted with fundamental obstacles only in the last. Advances in computer hardware and software have greatly reduced the earlier problems in these areas.

In spite of the progress that has been made in linguistic analysis, linguistic research has dealt primarily with syntactic analysis of individual sentences, and hardly at all with semantic problems and discourse analysis. As a result, current linguistic theory is inadequate for machine translation. For machine translation, semantic representations derived from syntactic structures in the source language must be associated with syntactic structures in the target language. See Katz-Postal, An Integrated Theory of Linguistic Descriptions, 1964, 166-172. To meet this problem, linguists must concern themselves with performance models and with semantic and discourse analysis. Moreover, comprehensive grammars do not yet exist for any language. Estimates of the availability of adequate grammars vary. Production of such grammars depends on the complexity of the model of language and on the research support provided. Some scholars participating in the Study suggest a date of five to ten years, a figure proposed also in a study carried out by Soviet specialists.

In view of the Peters-Ritchie results, it may be advisable to continue efforts with more restricted grammatical models which provide exact surface analysis based on syntactic and semantic features in the lexicon. Examples

are string analysis, the model used at the Linguistics Research Center, dependency theory of the Soviet type, and grammatical models whose transformational apparatus is more restricted than that of "standard" transformational grammars, for example, systems which use non-ordered or partially ordered transformations or equivalence transformations. Further, research in discourse analysis should be increased. Since the problems in machine translation are not the generation of coherent discourse but the carrying across of information, the achievement of translation would be considerably facilitated by such models. These problems may even be less pressing in actual practice because of the user reaction; that is, very often it may not be necessary for the system to represent all alternatives since the user will be able to provide the proper reading because of his access to information necessary for comprehension. Investigations on user-translation interaction should be carried out, especially in view of the highly divergent estimates of Zarechnak and Bar-Hillel. See also section 4, p. 46.

2. Like other technological applications, machine translation can be designed with various degrees of adequacy. The history of machine translation reflects this situation. The first attempts were primarily lexical. Syntactic analysis was then added. Currently semantic analysis is included for projected machine translation systems.

The improved understanding of language resulting from these progressively more comprehensive descriptions of language leads to improved translations. Translations based on semantic analysis will be correct when the information needed for disambiguation of a sentence is contained in that

sentence. When it is not, contextual and pragmatic information will be necessary.

3. Meaning is largely determined by the semantic readings of the lexical items in a sentence and the syntactic (semantic) relations between those items; these are presumably represented by the underlying structures of language. To arrive at the meanings of specific sentences, the underlying structure will have to be determined from the surface structure. In related languages, such as English and German, the relationships between surface and underlying structure are more similar than they are between less related languages like Russian and English or unrelated languages, such as English and Chinese. Accordingly, it will be simpler to devise translation systems for related languages. For the development of the technology of machine translation, systems designed for related languages are accordingly recommended at this time as an immediate goal. Medium-range goals (Russian-English) and long-range goals (Chinese-English) should also be planned.

4. The usefulness of translation depends on various factors: cost, timeliness, comprehensibility. In locations where imperfect, lexically-based machine translations are available, scientists have selected these over human translation when they could be made available the following day and human translations only after a week. In view of this situation, studies should be performed to measure the extent to which comprehensibility of a translation is dependent on the knowledge available to the actual user. Moreover, it should be noted that timeliness ranks high as a factor in translation. See also page 46.

Participants in the Study did not agree on what constitutes "high quality" translation. There is apparently no absolute standard. Rather, standards must be defined with reference to specific users and specific purposes.

Recommendations

1. On the basis of this Study it is recommended that support be made available for research in machine translation. The recommendation is made on the grounds that quality translation can be achieved in the near future. This recommendation agrees strikingly with conclusions reached in a study carried out in the Soviet Union.

Moreover, apart from attempts in information retrieval, machine translation is currently the only discipline which requires the study of problems beyond the sentence boundary. Because of the general lack of interest in these problems on the part of linguists, machine translation should be sponsored as an intellectual pursuit contributing to our knowledge of language.

2. For improved machine translation, research in the areas of descriptive linguistics, theoretical linguistics, comparative linguistics, stylistics, and evaluation of translation is necessary and should be supported.

2.1 Lexical research is necessary to determine the syntactic and semantic patterns of linguistic entities. Recent lexical research has indicated that entities such as verbs which have more than one meaning may have a particular meaning (1) only when they occur in specific syntactic environments whereas they have meaning (2) or further meanings when they occur in other specific environments. To illustrate the effect of only a trivially improved lexicon on translation, the report of an experiment conducted by Stachowitz in the spring of 1967 is appended.

2.2 Continued syntactic research, based on comprehensive lexical research, is essential. Fortunately a great deal of such research is being carried out, though eclectically, by linguists and their students in the normal course of their activities. Funding of such research should be increased, as well as linguistic study carried out in accordance with various approaches to language.

3. Theoretical research is essential, especially in view of the conclusions arrived at by Peters and Ritchie with reference to current transformational grammar. Various models based on differing grammatical assumption and/or using less powerful transformations should be investigated. Among these are the string analysis of Harris, and models like those of the Russian linguists which are based on dependency grammars. See Conclusions, p. 44.

Besides encouraging research in discourse analysis and production of coherent discourse, the possibility of establishing a research area intermediate between sentence (constituent) analysis and contextual analysis should be investigated. This investigation would be concerned with notions like those proposed by Fraser on "usage of sentences."

4. Contrastive studies dealing with the lexical and syntactic structures of two languages, and with the similarities of mapping these structures into semantic representations should be carried out. Such studies should also be concerned with the ways in which covert and overt semantic distinctions are expressed in these languages, including their overlap.

The overlap of lexical, syntactic and semantic ambiguity between two

languages should be studied from the point of view of carrying source language ambiguity over into the target language.

5. Descriptive stylistic studies on the incidence of lexical, syntactic and semantic ambiguities in scientific texts and their resolution by means of sentence immanent, context immanent and context external information should be carried out, as well as contrastive stylistic studies on scientific texts in contrast with literary texts. Such studies should aim to determine the types of syntactic structures used in various "styles" of language and possible divergences between them.

6. Explicit study should also be made of the kind of information available to the user which is necessary for the understanding of material that is mechanically translated. Such studies should seek to determine the amount of knowledge available from the surrounding text, as well as the amount of world knowledge necessary for the understanding of individual sentences. These investigations would be designed to determine the amount of information which must be provided to the machine so that the output is intelligible to a specialist or a general user.

7. Since the results of linguistic research will contribute to advances in machine translation, support is also recommended for research on problems in linguistics.

APPENDIX I

Conferences at the Linguistics Research Center

June 15 - 19, 1970

Martin Kay, W. P. Lehmann, Norman M. Martin, Jacob Mey, Stanley R. Petrick, Robert F. Simmons, Rolf Stachowitz, Donald E. Walker, Ladislav Zgusta, and the Linguistics Research Center staff.

September 30 - October 2, 1970

Emmon W. Bach, Charles Fillmore, Lauri J. Karttunen, W. P. Lehmann, John Lyons, Norman M. Martin, Jacob Mey, P. Stanley Peters, Robert F. Simmons, Rolf Stachowitz, and the Linguistics Research Center staff.

December 16 - 18, 1970

Emmon W. Bach, Lauri J. Karttunen, Robin and George Lakoff, W. P. Lehmann, James D. McCawley, Norman M. Martin, P. Stanley Peters, John R. Ross, Robert S. Simmons, Rolf Stachowitz, and the Linguistics Research Center staff.

January 11 - 15, 1971

Emmon W. Bach, Yehoshua Bar-Hillel, Paul L. Garvin, J. Bruce Fraser, Lauri J. Karttunen, Martin Kay, W. P. Lehmann, John Lyons, R. Ross Macdonald, Norman M. Martin, Z. L. Pankowicz, Eugene D. Pendergraft, P. Stanley Peters, Stanley R. Petrick, John R. Ross, Robert F. Simmons, Rolf Stachowitz, Rowena Swanson, Donald E. Walker, Terry A. Winograd, Michael Zarechnak, Ladislav Zgusta, and the Linguistics Research Center staff.

Visits of Individual Consultants

March - April 1970

Yehoshua Bar-Hillel

April 1970

Roman Jakobson

July 1970

J. Bruce Fraser

September 1970

Paul L. Garvin

February 1971

Harry H. Josselson

March 1971

Roman Jakobson

APPENDIX I I

Papers of participants

Supporting papers. As the statement on work under the Contract indicates, participants in the Study presented papers or participated in informal discussions. Some of the presentations reflect views of the authors which are published elsewhere. Participants were not pressed to provide papers for the final report. Further, the views which are expressed in the accompanying papers are strictly those of each author. Some of the authors have modified their statements after taking part in one or more conferences. But each paper included here was provided by its author and is given without modifications in the form provided originally by its author.

Most readers of the Report are probably acquainted with the authors of the appended papers. It has seemed unnecessary to provide introductions for such outstanding figures in linguistics and related disciplines. The appended bibliography, which is highly selective, will provide further access to the authors and their views.

Titles of Papers in Appendix I

Y. Bar-Hillel: Some Reflections on the Present Outlook for High-Quality Machine Translation

C= F. Fillmore: On a Fully Developed System of Linguistic Description

P. L Garvin: Operational Problems of Machine Translation: A Position Paper

L. Karttunen: The Logic of English Predicate Complement Constructions

J. Lyons: The Feasibility of High Quality Machine Translation

N. Martin: Philosophy of Language and the Feasibility of MT: A Position Paper

J. Mey Toward a Theory of Computational Linguistics

E. Pendergraft: Meaning Revisited

S. Petrick: Syntactic Analysis for Transformational Grammar

_____ Syntactic Analysis Requirements of Machine Translation

A. Stachowitz: Analysis of Es liegt eine grosse Anzahl von Elementen vor

R. Stachowitz: Lexical Features in Translation and Paraphrasing : an Experiment

_____ Requirements for Machine Translation: Problems, Solutions, Prospects

D. Walker The Current Status of Computer Hardware and Software as it Affects the Development of High Quality Machine Translation

L. Zgusta:Equivalents and Explanations in Bilingual Dictionaries

_____ The Shape of the Dictionary for Mechanical Translation Purposes

THEORETICAL STUDY EFFORT OF HIGH QUALITY TRANSLATION

Bibliography of Relevant Papers by Consultants and Major Contributors

BACH, EMMON

1964 *An Introduction to Transformational Grammars.* Holt, Rinehart & Winston.

1968 "Nouns and noun phrases", 91-122 in *Universals in Linguistic Theory.* Holt, Rinehart & Winston.

1971 "Syntax since *Aspects",*in *Report of the 22nd Annual Round Table Meeting,* Georgetown University Press.

BAR-HILLEL, YEHOSHUA

n.d. *Four Lectures on Algebraic Linguistics and Machine Translation.* ["Language and speech: theory vs. observation in linguistics", "The role of grammatical models in MT", "Syntactic Complexity", "Why machines won't learn to translate well".] Hebrew University, Jerusalem.

1955 "Idioms", 183-93 in *Machine Translation of Languages*, W.N. Locke and A.D. Booth, eds. , M.I.T. Press and John Wiley & Sons .

1959 *Report on the State of Machine Translation in the United States and Great Britain.* Hebrew University, Jerusalem.

1960 "The present status of automatic translation of languages", Appendix II in *Advances in Computers,* vol. 1, F.L. Alt, ed., Academic Press.

1961 "On formal properties of simple phrase structure grammars" [with M. Perles and E. Shamir]. *Zeitschrift für Phonetik, Sprachwissenschaft und Kommunikationsforschung* 14, ii: 144-72.

1963a "Is information retrieval approaching a crisis?" *American Documentation* 14,ii:95-98.

1963b "Remarks on Carnap's *Logical Syntax of Language",* 519-43 in *The Philosophy of Rudolf Carnap,* P.A. Schlipp, ed., Open Court Publishing Co.

BAR-HILLEL, cont'd

1965 "The outlook for computational semantics", I/1-14 in *Proceedings of the Conference on Computer-Related Semantic Analysis,* Wayne State University.

1966 "Do natural languages contain paradoxes?" *Studium Generale* 19,ix:391-97.

1967a "Dictionaries and meaning rules". *Foundations of Language 3:* 409-14.

1967b "Measures of syntactic complexity" [with A. Kasher and E. Shamir], 29-50 in *Machine Translation,* A.D. Booth, ed., North-Holland and John Wiley £ Sons.

1967c "Review of *The Structure of Language: Readings in the Philosophy of Language* (ed. by J.A. Fodor and J.J. Katz)". *Language* 43:526-50.

1968 "Universal semantics and philosophy of language: Quandaries and prospects", in *Substance and Structure of Language,* J. Puhvel, ed., University of California Press.

1969 "Formal logic and natural languages: A symposium". *Foundations of Language* 5:256-84.

1970 "Some reflections on the present outlook for high-quality machine translation (Position paper on MT in 1970)". LRC, University of Texas at Austin, xerox.

FILLMORE, CHARLES

1962 "Indirect object constructions in English and the ordering of transformations", 1-49 in *Project on Linguistic Analysis Report No. 1,* Ohio State Univ.

1963 "The position of embedding transformations in a grammar", 1-32 in *Project on Linguistic Analysis No.* 3, Ohio State University. [Also in *Word* 19:208-31.]

1965 "Entailment rules in a semantic theory". *Project on Linguistic Analysis Report No. 10,* Ohio State Univ.

1966a "Deictic categories in the semantics of *'come'". Foundations of language* 2:219-26.

FILLMORE, cont'd

1966b "A proposal concerning English prepositions", 19-33 in *Report of the 17th Annual Pound Table Meeting,* Georgetown University Press.

1966c "Toward a modern theory of case", 1-24 in *Project on Linguistic Analysis Report No. 13,* Ohio State Univ..

1967 "The grammar of HITTING and BREAKING", 9-29 in *Working Papers in Linguistics No. 1,* Ohio State University. [Also in *Studies in English Transformational Grammar,* R. Jacobs and P, Rosenbaum, eds., Ginn-Blaidsdell, 1970.]

1968a "The case for case", 1-88 in *Universals in Linguistic Theory,* E. Bach and R. Harms, eds., Holt, Rinehart & Winston .

1968b "Lexical entries for verbs". *Foundations of Language* 4:373-93.

1968c "Types of lexical information", 65-103 in *Working Papers in Linguistics No. 2,* Ohio State University. [Also to be in *Semantics: An Interdisciplinary Reader in Philosophy, Linguistics, Anthropology, and Psychology,* D. Steinberg and L. Jakobovits, eds., Cambridge University Press; and in *Proceedings of the Balatonszabadi Conference on Mathematical Linguistics,* F. Kiefer, ed. , D. Reidel.]

1969a "Review of *Componential Analysis of General Vocabulary: The Semantic Structure of a Set of Verbs in English, Hindi, and Japanese,* Part II, by Edward Herman Bendix". *General Linguistics* 9: 41-65.

1969b "Verbs of judging: An exercise in semantic description", 91-117 in *Papers in Linguistics* 1:5, Florida State Univ..

1970a "On a fully developed system of linguistic description". LRC, University of Texas at Austin, xerox.

1970b "On creativity", 2-19 in *Working Papers in Linguistics No. 6,* Ohio State University.

1970c "Subjects, speakers and roles", 32-62 in *Working Papers in Linguistics No. 4,* Ohio State University.

to appear *Studies in Linguistic Semantics,* as ed. with T. Langendoen.

FRASER, J. BRUCE

1965 *An Examination of the Verb-Particle Construction in English.* Ph.D. dissertation, M.I.T..

1966a "Some remarks on the verb particle combination in English", 45-6l in *Report of the l7th Annual Round Table Meeting,* Georgetown University Press.

1966b "Survey of automated language processing 1966", [with D.G. Bobrow & M.R. Quillian] Bolt, Beranek & Newman.

1967 "Some remarks on the characterization of syntactic transformation", mimeograph

1968a "A phonological rule tester", [with D.G. Bobrow]. *Communications of the ACM* 11:766-72.

1968b "Some remarks on the action nominalization in English", in *Readings in Transformational Grammar,* R. Jacobs and P. Rosenbaum, eds., Ginn-Blaisdell.

1969a "An analysis of concessive conditionals", 66-75 in *Papers from the Fifth Regional Meeting, Chicago Linguistic Society,* R.I. Binnick, et al., eds., Department of Linguistics, University of Chicago.

1969b "An augmented state transition network analysis procedure", 557-67 in *Proceedings of the International Joint Conference on Artificial Intelligence.*

1970a "Non-mind-changing verbs", 106-08 in *Language Research Report No. 2,* Language Research Foundation.

1970 "Idioms within a transformational grammar". *Foundations of Language* 6:22-42.

1970 "A reply to 'On declarative sentences'", 305-22 in *Mathematical Linguistics and Automatic Translation Report No. NSF-24,* Harvard Computation Laboratory.

1970 "Vice-versa". *Linguistic Inquiry* 1:277-278

to appear "An analysis of *even* in English", in *Proceedings of the Ohio State University Conference on Semantics.*

GARVIN, PAUL L.

1961 "Automatic linguistic analysis, a heuristic problem", 3-16 in *First International Conference on MT,* National Physical Laboratory, Teddington, England.

1962a "Computer participation in linguistic research". *Language* 38:385-89.

1962b "The impact of language data processing on linguistic analysis", *Inductive Methods in Semantic Analysis* Technical Note No. 3, Thompson Ramo Wooldridge, Inc..

1962c "A study of inductive method in syntax", *Inductive Methods in Semantic Analysis* Technical Note No. 1, Thompson Ramo Wooldridge, Inc..

1963 "Research in semantic analysis", *Inductive Methods in Semantic Analysis* Technical Note No. 6, Thompson Ramo Wooldridge, Inc..

1966a *Fulcrum Techniques for Chinese-English Machine Translation,* Final Report, Bunker-Ramo Corp..

1966b *Predication Typing-- A Pilot Study in Semantic Analysis,* [with J. Brewer and M. Mathiot], Bunker-Ramo Corp.

1966c "Some comments on algorithm and grammar in the automatic parsing of natural languages". *Mechanical Translation* 9:2-3.

1967a *Adaptation of Advanced Fulcrum Techniques to MT Production System (Russian-English)*, Bunker-Ramo Corp.

1967b "American Indian languages: Laboratory for linguistic methodology". *Foundations of Language* 3:257-60.

1967c "Heuristic syntax", *Computer-Aided Research in Machine Translation* Progress Report No. 14*,* Bunker-Ramo Corp.

1967d "Machine translation - fact or fancy?" *Datamation,* April.

1968 "The place of heuristics in the fulcrum approach to machine translation". *Lingua* 21:162-82.

1970 "Operational problems of machine translation: a position paper". LRC, University of Texas at Austin, xerox.

to appear *On Machine Translation.* Mouton.

JOSSELSON, HARRY H.

196la "Multiple meaning in machine translation", [with A. Janiotis] in *First International Conference on MT,* National Physical Laboratory, Teddington, England.

196lb "Research in machine translation", 160-72 in *Proceedings of the National Symposium on Machine Translation,* H.P. Edmundson, ed., Prentice-Hall.

1965 "Automatization of lexicography", paper presented at the 1965 International Conference on Computational Linguistics.

1967 "Lexicography and the computer", 1046-59 in *To Honor Roman Jakobson,* Mouton.

1968a "The lexicon: a matrix of lexemes and their properties", in *Proceedings of the Balatonszabadi Conference on Mathematical Linguistics,* F. Kiefer, ed. , D. Reidel.

1968b *Research in MT: Russian to English. Ten Year Summary Report.* Wayne State University.

1969a "The lexicon: a system of matrices of lexical units and their properties", paper presented at the International Conference on Computational Linguistics, Sänga-Såby, Sweden.

1969b "A linguistic interpretation of MT in the sixties", I/1-72 in *Research in Computer-Aided Translation, Russian → English,* 11th Annual Progress Report, Wayne State University.

1970 "The matrix as a concept for structuring the lexicon", Wayne State University, xerox.

KARTTUNEN, LAURI

.1967 "The identity of noun phrases". The RAND Corporation.

1968a "What do referential indices refer to?" The RAND Corp..

1968b "What makes definite noun phrases definite?" The RAND Corporation.

KARTTUNEN, cont'd

1969a "Discourse referents", Preprint No. 70, International Conference on Computational Linguistics, KVAL, Stockholm, Sweden.

1969b "Problems of reference In syntax", mimeograph.

1969c "Pronouns and variables", 108-16 in *Papers from the 5th Regional Meeting, Chicago Linguistic Society,* Department of Linguistics, University of Chicago.

1970a "The logic of English predicate complement constructions", LRC, University of Texas at Austin, xerox.

1970b "On the semantics of complement sentences", 328-39 in *Papers from the 6th Regional Meeting,* Chicago Linguistic Society.

to appear "Implicative verbs". *Language.*

KAY, MARTIN

1959 "The relevance of linguistics to mechanical translation", ML88/1-20 in *Essays on and in Machine Translation,* Cambridge Language Research Unit, England.

1960 *Mechanical Pidgin Translation* [with M. Masterman]. Cambridge Language Research Unit, Cambridge, England.

1967a "Experiments with a powerful parser", paper No. 10 In *Proc. Deuxième Conference International sur le Traitement Automatique des Langues,* Grenoble, France.

1967b "From semantics to syntax". The RAND Corporation.

1967c "Standards for encoding linguistic data in a natural language". *Computers and the Humanities* 1:170-77-

1969 "Computational competence and linguistic performance". The RAND Corporation.

1970a "The MIND system: The morphological-analysis program" [with G.R. Martins]. The RAND Corporation.

1970b "The MIND system: The structure of the semantic file" [with S.Y.W. Su]. The RAND Corporation.

1970c "Performance grammars". The RAND Corporation.

LAKOFF, GEORGE

1965 *On the Nature of Syntactic Irregularity,* Ph.D. dissertation, Indiana University. *Mathematical Linguistics and Automatic Translation Report No. NSF-16,* Harvard Computation Laboratory. [Also to appear as *Syntactic Irregularity,* Holt, Rinehart & Winston, 1970.]

1966a "A criterion for verb phrase constituency" [with J.R. Ross], II/1-11 in *Mathematical Linguistics and Automatic Translation Report No. NSF-17,* Harvard Computation Laboratory.

1966b "A note on Negation", III/1-16 in *Mathematical Linguistics and Automatic Translation Report No. NSF-17,* Harvard Computation Laboratory.

1966c "Phrasal conjunction and symmetric predicates" [with S. Peters], VI/1-49 in *Mathematical Linguistics and Automatic Translation Report No. NSF-17,* Harvard Computation Laboratory. [Also in *Modern Studies in English,* D. Reibel and S. Schane, eds., Prentice-Hall, 1969.]

1966d "Stative adjectives and verbs in English", I/1-16 in *Mathematical Linguistics and Automatic Translation Report No. NSF-17,* Harvard Computation Laboratory.

1968a "Deep and Surface grammar". Indiana University Linguistics Club, mimeograph.

1968b "Instrumental adverbs and the concept of deep structure". *Foundations of Language* 4:4-29.

1968c "Is deep structure necessary?" [with J.R. Ross]. Indiana University Linguistics Club, mimeograph.

1968d "Pronouns and reference", Indiana University Linguistics Club, mimeograph.

1968e "Some verbs of causation and change", in *Mathematical Linguistics and Automatic Translation Report No. NSF-20,* Harvard Computation Laboratory.

1969a "Baker's conjecture and natural logic", in *Phonetics Laboratory Notes* vol. 4, University of Michigan.

1969b "On derivational constraints", 117-39 in *Papers from the 5th Regional Meeting, Chicago Linguistics Society,* Department of Linguistics, University of Chicago.

LAKOFF, cont'd

1970a "Adverbs and modal operators", mimeograph.

1970b "Adverbs and opacity: a reply to Stalnaker", University of Michigan Phonetics Laboratory, mimeograph.

1970c "Counterparts, or the problem of reference in transformational grammar", 23-36 in *Mathematical Linguistics and Automatic Translation Report No. NSF-24.* Harvard Computation Laboratory.

1970d "Enough already" [with R. Lakoff], 19-20 in *Phonetics Laboratory Notes,* vol. 6, University of Michigan.

1970e "Global rules". *Language 4*6:627-39.

1970f *Linguistics and Natural Logic,* Studies in Generative Semantics No. 1, University of Michigan Phonetics Laboratory.

1970g "Natural logic and lexical decomposition", 340-62 in *Papers from the 6th Regional Meeting,* Chicago Linguistic Society.

1970h "A note on an arcane argument", 21-23 in *Phonetics Laboratory Notes,* vol. 6, University of Michigan.

1970i "Presuppositions and relative grammatica1ity", 51-68 in *Mathematical Linguistics and Automatic Translation Report NSF-24,* Harvard Computation Laboratory. [Also in *Philosophical Linguistics, Series 1,* W. Todd, ed., Great Expectations Pub. Co. (Evanston, Ill.), 1969.]

1970j "The role of deduction in grammar", 37-50 in *Mathematical Linguistics and Automatic Translation Report No. NSF-24,* Harvard Computation Laboratory.

to appear

"Gender agreement in French" [with D.M. Perlmutter]. *Linguistic Inquiry.*

"On generative semantics", in *Semantics: An Interdisciplinary Reader in Philosophy, Linguistics, Anthropology, and Psychology,* D. Steinberg and L. Jakobovits, eds., Cambridge University Press.

"Repartee". *Foundations of Language.*

"Some thoughts on transderivational constraint".

LAKOFF, ROBIN

1968 *Abstract Syntax and Latin Complementation.* MIT Press.

1969a "Some reasons why there can't be any *some-any* rule". *Language 4*5:608-15.

1969b "A syntactic argument for negative transportation", 140-47 in Papers *from the 5th Regional Meeting*, *Chicago Linguistic Society,* Department of Linguistics, University of Chicago.

1970a "If's, and's, and but's about conjunction", mimeograph, extended version of paper presented at the Ohio State University Spring Semantics Festival, 1969.

1970b "Tense and its relation to participants". *Language 4*6:838-49.

1970c "Enough already" [with G. Lakoff], 19-20 in *Phonetics Laboratory Notes,* vol. 6, University of Michigan.

to appear "Questionable answers and answerable questions", in *Papers in Linguistics in Honor of Henry and Renee Kahane,* B. Kachru, et al., eds.

LEHMANN, WINFRED P.

1964 "Linguistic theories underlying the work of various MT groups" [with M.L. Joynes]. *Linguistics* 8:62-71.

1965a "Interfaces of language", I/1-12 in *Proceedings of the Conference on Computer-Related Semantic Analysis,* Wayne State University.

1965b "Toward experimentation with language". *Foundations of Language* 1:237-48.

1968 *Research in Chinese Lexicography,* [with L.W. Tosh], LRC, University of Texas at Austin.

1969 *Research in Russian-English Machine Translation on Syntactic Level,* vols. 1 *& 2* [with L.W. Tosh, R.R. MacDonald, and M. Zarechnak]. LRC, University of Texas at Austin .

1970 *Research in German-English Machine Translation on Syntactic Level,* vols. 1 *&* 2 [with R. Stachowitz and L.W. Tosh]. LRC, University of Texas at Austin.

LYONS, JOHN

1963 *Structural Semantics.* Philological Society Publication 20, Blackwell.

1964 *Linguistic Transformational Analysis* [with J.P. Thorne]. RADC , Griffiss Air Force Base, New York.

1966a "Review of *An Integrated Theory of Linguistic Description,* J.J. Katz and P.M. Postal." *Journal of Linguistics* 2: 119-26.

1966b "Towards a 'notional' theory of the 'parts of speech'". *Journal of Linguistics* 2:209-36.

1967 "A note on possessive, existential, and locative sentences". *Foundations of Language* 3:390-96.

1970 "The feasibility of high-quality machine translation", LRC, University of Texas at Austin, xerox.

McCAWLEY, JAMES D.

1967 "The annotated respective", paper presented at the LSA Winter Meeting under the title "The respective downfalls of deep structure and autonomous syntax".

1968a "Concerning the base component of a transformational grammar". *Foundations of Language* 4:243-69.

1968b "Lexical insertion in a transformational grammar without deep structure", 71-80 in *Papers from the 4th Regional Meeting, Chicago Linguistic Society,* Department of Linguistics, University of Chicago.

1968c "Meaning and description of languages", *Kotoba no Uchii* 2, ix, x, xi.

1968d "Review of *Current Trends in Linguistics: Theoretical Foundations,* vol. 3, T.A. Sebeok, ed.". *Language 44:* 556-93.

1968e "The role of semantics in a grammar", 124-69 in *Universals in Linguistic Theory,* E. Bach and R. Harms, eds., Holt, Rinehart & Winston.

1970a "English as a VSO language", *Language* 46:286-99.

McCAWLEY, cont'd

1970b "Semantic representation", 227-48 in *Cognition: A Multiple View,* P. Garvin, ed., Spartan.

1970c "Where do noun phrases come from?" in *Readings in English Transformational Grammar,* R. Jacobs and P. Rosenbaum, eds. , Ginn-Blaisdell.

to appear "Tense and time reference in English", in *Proceedings of the First Ohio State Conference on Semantics,* C.J. Fillmore and D.T. Langendoen, eds., Holt, Rinehart & Winston.

MEY, JACOB

1965 "Pushdown stores and subscripts", paper presented at the International Conference on Computational Linguistics.

1968d "On the notion *to be* in Eskimo", 1-34 in *The Verb "Be" and Its Synonyms 2,* J.W. Verhaar, ed., D. Reidel.

1968b "Some shallow thoughts on deep structure". *Computing Reviews 9:429-34.*

1969 "Syntax or semantics: Some controversial issues in computational linguistics". *Norsk Tidsskrift for Sprogvidenskap* 23:59-70.

1970a "The cyclic character of Eskimo reflexivization". *Anthropological Linguistics.*

1970b "Norwegian generative grammar: Some antilegomena". *Norsk Tidsskrift for Sprogvidenskap* 24.

1970c "Possessives and transitive in Eskimo", *Journal of Linguistics* 6:47-56.

1970d "Toward a theory of computational linguistics", paper presented at the Annual Meeting of the Assn. for Computational Linguistics. LRC, University of Texas at Austin, xerox.

to appear "Comparatives in Eskimo".

PETERS, P. STANLEY, Jr.

1966a "A note on ordered phrase structure grammars", VIII/ 1-19 in *Mathematical Linguistics and Automatic Translation Report NSF-17,* Harvard Computation Laboratory.

1966b "Phrasal conjunction and symmetric predicates" [with G. Lakoff], VI/1-49 in *Mathematical Linguistics and Automatic Translation Report NSF-27,* Harvard Computation Laboratory. [Also 113-42 in *Modern Studies in English,* D. Reibel and S. Schane, eds., Prentice-Hall, 1969.]

1969a "Ambiguity, Completeness and restriction problems in the syntax-based approach to computational linguistics" [with R. Tabory], *Linguistics* 46:54-76.

1969b "A note on the universal base hypothesis" [with R.W. Ritchie]. *Journal of Linguistics* 5:151-52.

to appear "On the generative power of transformational grammars" [with R.W. Ritchie].

PETRICK, STANLEY R.

1965a "On the relative efficiencies of context-free grammar recognition" [with T.V. Griffiths]. USAF Cambridge Research Laboratories, Bedford, Mass..

1965b *A Recognition Procedure for Transformational Grammars.* Ph.D. dissertation, M.I.T.

1966 "A program for transformational syntactic analysis". USAF Cambridge Research Laboratories, Bedford, Mass..

1967 "Syntactic analysis" [with S.J. Keyser]. USAF Cambridge Research Laboratories, Bedford, Mass..

1969 "On coordination reduction and sentence analysis" [with P.M. Postal and P.S. Rosenbaum]. *Communications of the ACM* 12:223-33.

1971a "On the use of syntax-based translators for symbolic and algebraic manipulation", 224-37 in *Proceedings of the Second Symposium on Symbolic and Algebraic Manipulation.*

PETRICK, cont'd

1971b "Syntactic analysis for transformational grammars". LRC. University of Texas at Austin, xerox.

1971c "Syntactic analysis requirements of machine translation". LRC, University of Texas at Austin, xerox.

ROSS, JOHN ROBERT

1965 "Underlying structures in discourse" [with T.G. Bever], VIII/1-12 in *Proceedings of the Conference on Computer-Related Semantic Analysis,* Wayne State University .

1966a "A criterion for verb phrase constituency" [with G. Lakoff], V/1-3 in *Mathematical Linguistics and Automatic Translation Report NSF-17,* Harvard Computation Laboratory.

1966b "A proposed rule of tree-pruning", IV/1-8 in *Mathematical Linguistics and Automatic Translation Report NSF-17,* Harvard Computation Laboratory.

1966c "Relativization in extraposed clauses (A problem which evidence is presented that help is needed to solve", V/l-3 in *Mathematical Linguistics and Automatic Translation Report NSF-17,* Harvard Computation Laboratory.

1967a *Constraints on Variables in Syntax.* Ph.D. dissertation, M.I.T. Indiana University Linguistics Club, mimeograph, 1968.

1967b "On the cyclic nature of English pronominalization" , 1669-82 in *To Honor Roman Jakobson,* vol. 3, Mouton.

1968a "Is deep structure necessary?" [with G. Lakoff]. Indiana University Linguistics Club, mimeograph.

1968b "A note on 'command'", in *Mathematical Linguistics and Automatic Translation Report NSF-20,* Harvard Computation Laboratory.

1969a "Adjectives as noun phrases", 352-60 in *Modern Studies in English,* D. Reibel and S. Schane, eds., Prentice-Hall.

ROSS, cont'd

1969b "Guess who?" 252-86 in *Papers from the 5th Regional Meeting, Chicago Linguistic Society*, Department of Linguistics, University of Chicago.

1970 "On declarative sentences", in *Readings in English Transformational Grammar,* R. Jacobs and P. Rosenbaum, eds., Ginn-Blaisdell.

to appear "Gapping and the order of constituents", in *Proceedings of the 10th International Congress of Linguists,* Bucharest.

SIMMONS, ROBERT F.

n.d. "Natural language processing and the time-shared computer", 217-27 in *Toward a National Information System.*

1963a "A computational approach to grammatical coding of English words" [with S. Klein]. *Journal of the ACM* 10:334-37.

1963b "Co-occurrence and dependency logic for answering English questions" [with S. Klein and K. McConlogue]. System Development Corp..

1963c "Design for a language processing laboratory". System Development Corp.

1963d "Maximum-depth indexing for computer retrieval of English language data" [with K.L. McConlogue]. *American Documentation* 14:68-73.

1963e "Syntactic dependence and computer generation of coherent discourse". *Machine Translation* 7:50-61.

1963f "Synthetic language behavior". *Data Processing for Management* 5,xii:11-18.

1964a "Indexing and dependency logic for answering English questions" [with S. Klein and K. McConlogue]. *American Documentation* 15: 196-204.

1964b "Synthex 1964: A progress report and a research plan." System Development Corp.

SIMMONS, cont'd

1965a "Analyzing English syntax with a pattern-learning parser" [with K. McConlogue]. *Communications of the ACM* 5:687-98.

1965b "Natural-language processing by computers-- 1966". System Development Corp..

1966a "An approach toward answering English questions from text" [with J.F. Burger, and R.E. Long], 357-63 in *Proceedings of the 1966 Fall Joint Computer Conference,* Thompson Books.

1966b "Automated language processing", 137-69 in *Annual Review of Information Science and Technology,* C. Cuarda, ed., Interscience.

1966c "Storage and retrieval of aspects of meaning in directed graph structures". *Communications of the ACM* 9:211-15.

1967a "Answering English questions by computer", 253-89 in *Automated Language Processing,* H. Borko, ed., John Wiley & Sons. [Also in *Communications of the ACM* 8(1965):53-70.]

1967b "An interactive system for computing dependencies, phrase structures and kernels" [with J.F. Burger and R.E. Long]. System Development Corp..

1968a "A deductive question answerer for natural-language inference" [with R.M. Schwarcz and J.F. Burger]. System Development Corp..

1968b "A semantic analyzer for English sentences" [with J.F. Burger]. *Machine Translation* 11:1-13.

1969 "A computational model of verbal understanding" [with J.F. Burger and R.M. Schwarcz], 441-56 in *Proceedings of the 1968 Fall Joint Computer Conference,* Thompson.

1970a "Generating English discourse from semantic networks" [with J. Slocum]. Computer Assisted Instruction Laboratory, University of Texas at Austin.

1970b "Natural language question-answering systems". *Communications of the ACM* 13:15-31.

SIMMONS, cont'd

1970c "Some relations between predicate calculus and semantic net representations of discourse". Computer Sciences Dept., University of Texas at Austin, mimeograph.

1970d "Some semantic structures for representing English meanings". Computer Sciences Dept., University of Texas at Austin, mimeograph.

STACHOWITZ, ANNETTE

1971 "Analysis of *'Es liegt eine grosse Anzahl von Elementen vor.*'" LRC, University of Texas at Austin, xerox.

STACHOWITZ, ROLF

1969 *On the definition of the term "discontinuous constituents" in context-free phrase structure grammar.* Ph.D. dissertation, University of Texas at Austin.

1970a "A model for the recognition and production of synonymous expressions with different deep structures." LRC, University of Texas at Austin, xerox.

1970b *Research in German-English Machine Translation on Syntactic Level* vols. 1 & 2 [with W.P. Lehmann and L.W. Tosh]. LRC, University of Texas at Austin.

SWANSON, ROWENA

1967 *MOVE THE INFORMATION. . . . A Kind of Missionary Spirit.* USAF Office of Aerospace Research, Arlington, *Va.*

1970 "Trends in information handling in the United States". USAF Office of Scientific Research. Arlington, Va..

WALKER, DONALD E.

1962 "The structure of languages for man and computer: problems in formalization", paper presented at the First Congress of the Information System Sciences. The MITRE Corp.

1965a "English preprocessor manual". The MITRE Corp..

1965b "The MITRE syntactic analysis procedure for transformational grammars" [with A.M. Zwicky, J. Friedman, and B.C. Hall], 317-26 in *Proceedings of the 1965 Fall Joint Computer Conference,* Spartan.

1966 "Recent developments in the MITRE syntactic analysis procedure" [with P.G. Chapin, M.L. Geis, and L.N. Gross]. The MITRE Corp.

1967a "On-line text processing: Introduction and overview". The MITRE Corp.

1967b "SAFARI, an on-line text-processing system", 144-47. in *Proceedings of the American Documentation Institute.*

1969a "Computational linguistic techniques in an on-line system for textual analysis". The MITRE Corp..

1969b "On-line computer aids for research in linguistics" [with L.N. Gross], 1531-36 in *Information Processing 68,* North-Holland.

1970 "The current status of computer hardware and software as it affects the development of high quality machine translation". LRC , University of Texas at Austin, xerox.

ZGUSTA, LADISLAV

1970a "Equivalents and explanations in bilingual dictionaries", paper presented at the Modern Language Assn. lexicographical conference, Columbus, Ohio. LRC, University of Texas at Austin, xerox.

1970b "The shape of the dictionary for mechanical translation purposes", LRC, University of Texas at Austin, xerox.

to appear *Manual of Lexicography.*