SUMMARY OF THE PROCEEDINGS
OF THE WAYNE STATE UNIVERSITY CONFERENCE
OF FEDERALLY SPONSORED
MACHINE TRANSLATION WORKERS

Held At

Princeton Inn, Princeton, New Jersey

July 18-22, 1960

Sponsored Jointly By:

National Science Foundation

and

Office of Naval Research, Information Systems Branch

The idea of holding a meeting, devoted to a consideration of immediate problems confronting machine translation workers, had occurred to a number of us as early as the Fall of 1959. The idea gained impetus and grew in the course of informal discussions among some of the participants at the Machine Translation Conference sponsored by UCLA, February 2-5 of this year. It was then felt that what the machine translation field needed, above all, was for a number of investigators to get together in order to take inventory of the actual work being carried out in the field, and to lay a foundation for possible future exchange of concrete results. It was thought necessary to make the number of participants as small as possible in order to achieve positive results. It was thus decided to limit the meeting to representatives of federally sponsored machine translation groups only.

Accordingly, a request for financial assistance to defray the expenses of this meeting was directed to the National Science Foundation and the Information Systems Branch of the Office of Naval Research, both of whom granted generous support to the meeting. An organizing committee for the meeting, consisting of Harry H. Josselson, Sydney M. Lamb, and Victor H. Yngve, was formed. This committee held a meeting in Detroit in the middle of June, at which time the agenda for the meeting was drawn up.

The conference was held at Princeton Inn, Princeton, New Jersey, July 18-22, and was devoted largely to technical discussions of comparing and exchanging the work of the various machine translation groups up to date. Participants were asked to come to the meeting prepared with handouts, if possible, to explain in detail the codes, formats, decks and other pertinent material, that are available for distribution. The proceedings were summarized and are being distributed in the hereinclosed form.

The committee is deeply grateful to all those, both workers and sponsors, who took the trouble to attend the meeting, and whose presence and dedicated efforts made this session the success that it was, at least in the opinion of those who participated, as exemplified by the following comments:

> "I should like to express, somehow, in a public way, (our) appreciation of the meeting, which, in (our) opinion, was the best since the 1956 meeting at M.I.T., which indeed it resembled."
>
> "Congratulations on a very successful meeting."
>
> "I want to thank you for your hospitality at Princeton. The Wayne Conference was undoubtedly the most enjoyable that I have attended, and doubly so because of the opportunities it afforded for exchanges of information."

Thanks are also due of course to the National Science Foundation and the Information Systems Branch of the Office of Naval Research, for their generous support, financial and moral, as well as all the other sponsors who contributed so actively and fruitfully to the success of the meeting.

It is hoped that this modest beginning augurs well for future profitable exchanges of views and information in the machine translation field in this country and abroad.

September 1960

The Organizing Committee,

Harry H. Josselson

Sydney M. Lamb

Victor H. Yngve

CONTENTS PAGE

WORKING MT CONFERENCE PARTICIPANTS
Princeton Inn
Princeton, New Jersey
18-22 July 1960

Dr. Franz Alt
Applied Mathematics Division
National Bureau of Standards
Washington 25, D. C.

Dr. A. P. R. Brown
Machine Translation Research
Georgetown University
1715 Massachusetts Avenue
Washington, D. C.

Dr. Silvio Ceccato
Via Generale Arimondl 13
Milan, Italy

Dr. Douglas Ellson
Department of Psychology
University of Indiana
Bloomington, Indiana

Mrs. Joan Frye
Machine Language Translation Study
Box 7980, University Station
The University of Texas
Austin 12, Texas

Dr. Gordon Goldstein
Information Systems Branch
Department of the Navy
Office of Naval Research
Washington 25, D. C.

Dr. Kenneth Harper
Mathematical Analysis Department
The RAM) Corporation
Santa Monica, California

Dr. Paul W. Howerton
Central Intelligence Agency
2430 E. Street, N. W.
Washington 25, D. C.

Miss Amelia Janiotis
Department of Slavic and Eastern Languages
Machine Translation
Wayne State University
Detroit 2, Michigan

Mr. C. Douglas Johnson
Computation Center
University of California
Berkeley 4, California

Dr. Harry H. Josselson, Chairman
Department of Slavic and Eastern Languages
Wayne State University
Detroit 2, Michigan

Professor Sydney Lamb
Computation Center
University of California
Berkeley 4, California

Dr. David Lieberman
Massachusetts Institute of Technology
Room 20 D - 102
Cambridge 39, Massachusetts

Dr. Dean Lytle
College of Engineering
University of Washington
Seattle, Washington

Mrs. Margaret Masterman Braithwaite
Cambridge Language Research Unit
20 Millington Road
Cambridge, England

Mr. Scott A. McGall
U. S. Army Signal Research
and Development Laboratory
19 Madison Avenue
Long Branch, New Jersey

Dr. Leroy Meyers
Applied Mathematics Division
National Bureau of Standards
Washington 25, D.C.

Dr. H.W. Swarm
College of Engineering
University of Washington
Seattle, Washington

Mr. Roger Needham
Cambridge Language Research Unit
20 Millington Road
Cambridge, England

Dr. Victor H. Yngve
Massachusetts Institute of Technology
Room 20 D - 102
Cambridge 39, Massachusetts

Mr. Eugene Pendergraft
Machine Language Translation Study
Box 7980, University Station
The University of Texas
Austin 12, Texas

Dr. Marshall Yovits
Information Systems Branch
Department of the Navy
Office of Naval Research
Washington 25, D.C.

Mr. Oskar Reinson
Rome Air Development Center
Griffiss Air Force Base
Rome,New York

Mr. Michael Zarechnak
Machine Translation Research
Georgetown University,
1715 Massachusetts Avenue
Washington, D.C.

Dr. Richard See
Program for Documentation Research
National Science Foundation
Washington 25, D.C.

Mr. Ted Ziehe
Mathematical Analysis Department
The RAND Corporation
Santa Monica, California

Agenda for Princeton Meeting

Monday

A.M.

10:30-12:00 Informal discussion of common problems.

P.M.

2:00-4:00 Opening Session: Administrative Matters. (for Sponsors only)
Chairman: Richard See

Tuesday

A.M. Chairman: Sydney Lamb
9:00-10:15 Georgetown University
10:45-12:00 Massachusetts Institute of Technology

P.M. Chairman: Harry Josselson
2:00-3:15 University of California at Berkeley
3:45-5:00 Cambridge Language Research Unit

Wednesday

A.M. Chairman: Victor Yngve
9:00-10:15 Wayne State University
10:45-12:00 Centro di Cibemetica di Milano

P.M. Chairman: Harry Josselson
2:00-3:15 RAND Corp.
3:45-5:00 Meeting with Sponsors

Thursday

A.M. Chairman: Sydney Lamb
9:00-10:15 National Bureau of Standards
10:45-12:00 University of Washington
University of Texas
University of Indiana

P.M.

2:00-5:00 General Topic: Discussion of Possibilities of Coordinating Formats for Future Work

Chairman: Sydney Lamb
Topic: Formats for the Future
Possibilities:
a) semantic coding
b) programming languages
c) others

Friday

A.M.

9:00-12:00 General Topic: Discussion of Possibilities of Coordinating Formats for Future Work

Chairman: Victor Yngve

Topic: Possibilities for the Interconvertibility of Present Materials, Codes, and Formats.

a) dictionary (including grammar coding)
b) text
c) programs (including dictionary lookup and syntax)
d) others

GEORGETOWN UNIVERSITY PRESENTATION Tuesday, 19 July, 9:00-10:15 a.m.

ZARECHNAK

Mr. Zarechnak opened the session with general comments concerning the problems of dictionary storage. He added that storage would include only those features which are constant, and stems without endings.

Re explained that, serving as a basic approach to MT, an order of precedence would first attempt to solve the problem of a dictionary, and second, the problem of syntactic analysis. Since the sentence is the vehicle which carries the message from the source language to the target language, sentence structure is the level on which analysis of the text should begin. (Note: The work of this project has dealt with chemical texts.)

Mr. Zarechnak further stated that a division of functional analysis could be made onto three distinct levels: first, analysis of the morphological structure of the particular word, independent of its neighboring words; second, syntagmatic or continuous function analysis; and third, syntactic or discontinuous function analysis, e.g., nesting. A fourth level of analysis will necessarily be semantic analysis. In fine, the four levels of analysis are: morphological (word without neighbor), syntagmatic (continuous function), syntactic (discontinuous function), and semantic (e.g., Rain-refreshed forest).

Mr. Zarechnak proceeded to point out that there were obviously many immediate problems in the field of MT, such as subject-predicate relationship, and, in general, those problems encountered in the process of carrying a message from the source to the target language. He also announced that MT must have a series of language sciences to meet and solve the abovementioned problems. He concluded by acknowledging the immensity of the field and the vast contributions that are yet to be made.

GEORGETOWN UNIVERSITY PRESENTATION

A.F.R. BROWN

Dr. Brown had nothing he felt he might offer in the way of linguistic information, in view of the fact that he has spent the past fourteen months concentrating on questions of programming only. A significant product of this fourteen month period is Dr. Brown's "Simulated Linguistic Computer".

Dr. Brown presented his handout A Symbolic Language for Programming the Simulated Linguistic Computer, and taking the word 'haut' as an example (Dr. Brown's work has dealt exclusively with French), he discussed and graphically demonstrated an 'up-dating' procedure.

A brief question-answer discussion period followed. A question of major concern involved the quantity of text that should be required in order to form positive conclusions. It was generally agreed that it would depend upon both the amount of attention directed toward the text, and the extent to which one would rigidly adhere to established categories. There was also general agreement with Mrs. Masterman's comment that it was essential for the message to be preserved, that one could not determine what had been missed in translation by simply reading output.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY PRESENTATION Tuesday, 19 July, 10:45-12:00 a.m

LIEBERMAN

Dr. Lieberman presented a handout concerning a search routine, prepared by Ken Knowlton. Dr. Lieberman offered some general statistical information about the routine. He said that the input for this routine must be punched in a specific manner, which is worked out by the U.S. Patent Office and M.I.T. He further explained that each occurrence is given an integral number of machine words and that as many as one hundred items could be searched for at one time.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY PRESENTATION

At the blackboard, Dr. Lieberman drew a representative flowchart and offered information and explanations of what happened in the actual search. He demonstrated that several requests might be satisfied by one text sequence. Search time for scanning 200,000 words of text is about ten minutes plus 0.2 seconds for each encounter if context is to be printed out.

The source material which was used included: 100,000 words each of 1) Associated Press Material, 2) German Newspapers, 3) Patent Office Material.

Some general discussion of the handout text ensued.

YNGVE

Dr. Yngve initially offered some general comments about COMIT. He added that the program was to be distributed through SHARE. He then presented his approach, with particular emphasis centered around the 'depth phenomenon' and subsequent phrase structure. He treated related questions such as: how such memory is needed for specific procedures; e.g., expansion of the sentence into (a) subject and (b) predicate.

He proceeded with the presentation, offering a definition of the 'depth of a node' as being "the number of right branches required to go from that node back to the top". In estimating the size of a temporary memory, he suggested that a memory of about seven items is needed for producing English. He added that one result of the depth phenomenon is that we now have a definite reason for explaining why some sentences are awkward.

Dr. Yngve then discussed unordered phrase-structure rules, adding that a grammar of this kind can be constructed, as is implied in the M.I.T. handout. He then presented some sample output from a COMIT program, designed to generate

MASSACHUSETTS INSTITUTE OF TECHNOLOGY PRESENTATION

sentences at random. He explained that the program was text-oriented, that he had used a childrens book, Engineer Small, which, with its forty word vocabulary, was understandably limited. The product result is output without initial input.

Dr. Yngve concluded with an invitation for open discussion. He also invited all interested conferees to gather in the conference room, Tuesday evening at 8:30, for an informal discussion and explanation of COMIT.

BERKELEY PRESENTATION Tuesday, 19 July, 2:00-3:l5 p.m.

LAMB

Professor Lamb began his presentation by taking an arbitrary and discontinuous Russian sentence plus a good translation of it. Placing it on the blackboard, he proceeded to work out a lexeme by lexeme assignment of the translation. He offered as his main topic for discussion: the idea of using lexemes in a translation system. This topic, he added, could be placed under the heading of "helpful concepts and ideas", as had been suggested in the invitation letters in the way of recommendations for conference presentations.

Professor Lamb next invited the conference participants to look at the Berkeley handout material that he had distributed earlier, as he proceeded to discuss the individual items. First, he explained his Diagram of the Structure of a Translation System and discussed the Types of Relationships Between Levels. He made a point of emphasizing the fact that the advantage of this linguistic system was its simplicity.

To continue, he posed this question: what statements can be made about a linguistic system? And then, to answer his own question, he stated that there are two and only two types of statements needed:

1) How any item is related to other items on adjoining levels.

2) How items are related to items on the same level.

The purpose of statements of the second type is to describe the patterns of arrangement which occur. A complete description of arrangements would include a list of distribution classes of lexemes, and a list of constructions.

He added various other points of information such as the concept of the metataxeme, i.e., feature of arrangement in the target language, and also the fact that lexemes are commonly combinations of morphemes and parts of words.

At this point, Professor Lamb interjected some historical information concerning linguistic systems and two traditions in linguistic approach. He discussed the Hindu grammarians and their work with Sanskrit, a tradition which was continued by Bloomfield and others in the field, resulting in what is sometimes called the Item-Arrangement (IA) system. Secondly, ha spoke briefly about the Latin grammarians and their work, using the Word-Paradigm (WP) system, in a tradition which has been continued by most language teachers in this country. And finally he commented on IP - Item-Process, making reference to the contributions of Boas and Sapir. He pointed out that structural linguists have in recent years generally regarded the IA system as superior to others, to the extent that they sometimes even forget about the existence of the WP approach. Yet the latter, he stated, is being used almost universally by workers in machine translation.

BERKELEY PRESENTATION

After he offered a definition of a lexeme as "the basic unit of the dictionary or lexicon", Professor Lamb made some observations on lexemes in general, and then, turning back to the handout, shifted the discussion to nonce forms (forms coined as combinations of items), and related material on segmentation.

Professor Lamb talked about the productivity of Russian suffixes, as he presented his handout on Derivational Suffixes. He introduced the second Berkeley conferee, C. Douglas Johnson, who presented material along with the handout A List of Derivational Suffixes Considered for Segmentation. Immediately thereafter, Professor Lamb submitted comments on productivity in the source language as the main criterion for determining the proper degree of segmentation. He added that combinations which are complicated are not segmented.

The remaining time was spent in active open discussion.

CAMBRIDGE LANGUAGE RESEARCH UNIT PRESENTATION Tuesday, 19 July, 3:45-5:00 p.m.

MASTERMAN

Margaret Masterman (Mrs. Braithwaite) presented four CLRU items to the Meeting:

1) A flexible procedure for punched-card distribution (from a forthcoming CLRU Report), by M. Kay and T.R. McKinnon Wood.
2) Mechanical Pidgin Translation, a handout, of some 175 pages, reporting on a CLRU inquiry on the "language" produced by word-for-word M.T., of the kind at present being carried out by I.B.M. Research.

CAMBRIDGE LANGUAGE RESEARCH UNIT PRESENTATION

3) The resolution of Semantic Translation problems with the aid of a thesaurus, On this she asked the Meeting's leave to speak informally, and at some later time.

4) Dr. Parkers-Rhodes' Syntax-Finding Program. She introduced Mr. R.M. Needham to speak on this.

In passing, however, she stressed the value of cooperative exchange among the different research projects. She expressed her belief that some groups had assumed patterns of general research, while others had concentrated on particular aspects only. She anticipated genuine contributions from exchange between the particular - and generally - oriented groups.

NEEDHAM

Mr. Needham first presented the CLRU Bracketing Program. He explained that they had found it was possible to discover dependency and government relationships in text material, using unexpectedly simple syntactic coding. He added that with the blocking routine, titivation (homograph resolution) is carried on alternatively with bracketing, rather than doing everything in two separate stages.

Mr. Needham also described Parker-Rhodes' Rule for Bracketing, and thereafter, proceeded to offer a graphic example of how a dictionary entry is made. He also presented some CLRU handout material in conjunction with his demonstration.

In summation, he added that the system could be adapted to another language, the only changes made being in the dictionary and titivation routines. Mr. Needham offered to answer any questions from the floor.

CAMBRIDGE LANGUAGE RESEARCH UNIT PRESENTATION

Two questions receiving primary attention in the following open discussion period were concerned with scanning technique and the order of precedence to be taken regarding volume of data and awkward cases. It was generally agreed that scanning should be done back and forth, and there remained some mixed feeling about whether or not volumes of data should be taken first, as opposed to the immediate analysis of awkward examples.

WAYNE STATE UNIVERSITY PRESENTATION Wednesday, 20 July, 9:00-10:15 a.m.

JOSSELSON

Dr. Josselson's presentation consisted of a detailed description of the grammar coding scheme which the Wayne group is presently using. He discussed the 'part of speech' categories and the differences between the present and traditional grammar classes.

The coding sheet contains information to be used in the process of making translation decisions on both syntactic and semantic levels. In many instances a bit of information in the grammar code applies to a set of words, and a list of words in this set was included in the instructions. Dr. Josselson noted that the lists were in many cases merely a beginning, and that they could and would be expanded. He pointed out that one task for MT investigators is to seek and record examples of linguistic phenomena. He added that the questions asked in the coding format will change on the basis of further syntactic investigation; new categories will appear, and others may turn out to be unnecessary.

WAYNE STATE UNIVERSITY PRESENTATION

JANIOTIS

Miss Janiotis briefly discussed a 709 interpretive subroutine for machine translation problems (a description and flowcharts appear in the Wayne handout). She answered several questions and then proceeded to discuss nominal, prepositional, and governing modifier blocking routines, as they appear in the Wayne handout. She noted that the blocking routines were similar to that which was offered earlier by Mr. Needham of CLRU, under the title of Bracketing.

Miss Janiotis elaborated on the Nominal Blocking Routine and the remaining time was spent in open discussion of both Dr. Josselson's and Miss Janiotis' presentations.

CENTRO DI CIBERNETICA DI MILANO PRESENTATION Wednesday, 20 July, 10:45-12:00 a.m.

CECCATO

Dr. Ceccato prefaced his presentation with an announcement of his three hundred page report which is going through final proofing, and which he offered to mail to all interested conference participants as soon as it is published. He also wished to point out that the work he and his staff are presently doing is neither dictionary-grammar nor syntax oriented, so much as it is directed toward semantic analysis.

Dr. Ceccato continued by explaining more specifically that his group is trying to produce what is virtually a thinking machine which will simulate the processes of the human mind. According to Dr. Ceccato, the processes of the human mind involve a series of prescribed and fixed

operations; moreover, the problem at hand could be reduced to two questions that confront the investigator: (1) What is the structure of our thought? and (2) How are we to put a link between our language and our thought?

He attempted to clarify his hypothesis further by drawing several diagrams on the blackboard, first presenting the thought process as a product of what he termed the "correlator" and the "correlation", and second, drawing several examples from simple English and Italian phrases, and analyzing them in terms of his thought process box diagram.

Dr. Ceccato continued to elaborate on the function of the "correlator", adding parenthetically, that while some languages relied upon form (declension and inflection), others relied upon order (context). But he explained that it was not the language that changed; rather, it was the thought, and for us the correlation is done by the machine.

After some remarks about his two levels of language, i.e., the language itself, and those things that operate the language, Dr. Ceccato invited the group to gather around him as he presented and explained graphical data, including coding material and charts.

RAND CORPORATION PRESENTATION Wednesday, 20 July, 2:00-3:l5 p.m.

ZIEHE

Mr. Ziehe began the session by discussing the RAND handout Available RAND Linguistic Data. In discussing the text and dictionary he defined:

(a) an occurrence as an instance of a form in text

(b) a form as a unique sequence of alphabetic characters that is preceded and followed in text by either spaces and/or punctuation

(c) a word as the collection of forms that constitute a paradigm

RAND CORPORATION PRESENTATION

Mr. Ziehe also discussed the information carried by "special codes" - for equivalent inflection and idiom participation.

A tape dictionary is being developed at RAND. The entry for each form will consist of a number of variable length items. The number of items can be easily increased or decreased. He then discussed syntactic rules embodied in the RAND Dependency Table. He noted that the constructions not covered by the table are low frequency occurrences.

HARPER

Dr. Harper briefly mentioned recent publications describing the RAND sentence-structure determination system and the results of syntactic analysis. A handout showed the variety of existing analytic reports in which are recorded the syntactic combinations that have occurred in text processed to date. These reports are still being used for retrieval and coding of syntactic information. An example given was the identification of modals that are dependent upon the infinitive, and an indication of their relative position.

Dr. Harper then branched into a discussion of distributional semantics, and its relation to MT. In this approach, structurally related items are considered in terms of individual words; distributional classes may be formed on the basis of a) morphology, b) a priori considerations, or c) syntactic relationship to other distributional classes. Large samples of text will be required for the building of these classes.

The presentation was followed by open discussion and questions addressed to Mr. Ziehe and Dr. Harper.

MEETING WITH THE SPONSORS Wednesday, 20 July, 3:45-5:00 p.m.

1. At the second meeting of the Interagency Committee on Mechanical Translation Research, which was held Monday, 18 July, two guide-line statements were adopted concerning mechanical translation research.

   a) Statement on reporting:

      "In the field of research on scientific information problems, we consider it essential to progress in the field that full accounts of all aspects of such research work be made available promptly. These accounts should include the actual procedures developed and tested, all relevant data (except where the data are so voluminous that their reproduction and distribution would present difficulties), and the results achieved. The work should be reported in such a way that another investigator could, if he wished, confirm the results himself. Such full accounts will frequently be too voluminous for journal publication and should therefore be issued promptly in report form and, if appropriate, summarized for journal publication."

   b) Statement on meetings:

      "The Committee encourages productive meetings in the field of mechanical translation research and feels that it is appropriate to take an active part in the planning of such meetings in the future. This should be done, if possible, with the cooperation of any profes-sional society which may exist. The committee is in favor of

three types of meetings:

1) working-group meetings

2) large, open, national meetings

3) international meetings

The Committee also encourages visits among various groups."

In the statement on meetings, "professional society" refers to any society or societies in the field of mechanical translation research which may be formed at some future date.

The above two statements were read at the meeting with the sponsors by Richard See, Chairman of the Interagency Committee on Mechanical Translation Research.

2. The statement on reporting inspired considerable general discussion about the problem of exchanging information and material among the several projects. The problem of distribution of bulky reports and data was discussed, and microfilming suggested as one solution. It was pointed out that intelligibility may be an even more serious problem than bulkiness in some cases. It was agreed that there should be more coordination.

3. Another question concerned the frequency of progress reports. Dr. Yovits pointed out that quarterly progress reports are in some cases intended for the sponsor only.

4. The subject of meetings was discussed. Dr. Yovits pointed out that meetings like the present conference should be encouraged, but not indiscriminately; moreover, there should be adequate notice. It was noted that those actually involved in mechanical translation

research were in a better position to decide what kinds of meetings should be held and when, than those outside the field.

5. It was also suggested that communication with the Meetings Committee under Leon Dostert is necessary and desirable in connection with all meetings relating to mechanical translation which come to the attention of workers in the field.
6. It is the responsibility of the professional research personnel to arrange meetings, but the sponsors also have a legitimate concern with arrangements for meetings.

NATIONAL BUREAU OF STANDARDS PRESENTATION Thursday, 21 July, 9:00-10:15 a.m.

ALT

Dr. Alt began his presentation with general commentary on the NBS' approach. He said that their project has not been working on a dictionary, which they feel would have to be re-worked before long and hence would prove to be wasteful. They have concentrated on grammar which they feel should be worked out first, although semantics is also an important consideration. Economy has been stressed in their approach, and they aim at translation rather than research for its own sake. They consider themselves an intermediate-range project and favor conventional grammar. The basis of their method is expounded in an NBS Report by Ida Rhodes, distributed about a year ago, which will appear in a coining issue of MT

According to Dr. Alt, the NBS machine code is divided into two major parts. The first part is further sub-divided into (a) dictionary look-up and (b) the morphology of individual words. The second part of the machine code is sub-divided into (a) profile, (b) primary syntax, and (c) English.

The second part considers complete sentences only.

Dr. Alt then elaborated on Part I, first noting that morphological approaches of most research groups are equivalent. He stated that their process of making a very compact stem dictionary means more machine time and additional morphological analysis but saves storage space. In their proposed dictionary, storage is arranged by roots. The machine stores lists of prefixes and suffixes. A list of endings indicates those parts of speech and inflectional paradigms with which each ending can go.

Dr. Alt explained Part II of the machine code, as he directed the attention of the group to the NBS handout material. He said that the general procedure was iterative and involved a series of predictions, divided among glossary predictions, grammar predictions, and a few others. Each prediction is assigned an urgency number, and when the prediction is satisfied, it is erased. Unsatisfied predictions are erased if their urgency is low. Those possessing high urgency numbers are kept until the end of the iteration and serve as criteria of the goodness of the translation. Concerning Part II (a), the profile, he explained this procedure as a preliminary determination of the boundaries of the clauses and phrases of each sentence. This knowledge is imperative for syntactic analysis, since predictions can be made only within individual clauses.

Dr. Alt's final point, in discussing the handout, concerned the concept of "Hindsight", which is a part of the general procedure serving as follow-up for the prediction. There were four kinds of hindsight:

(i) no match between the Foresights and the morphology of an occurrence (symbolized by 'H_0')

NATIONAL BUREAU OF STANDARDS PRESENTATION

(2) too many such matches (symbolized by 'H_2')

(3) doubtful choices (symbolized by 'H_1')

(4) any morphological alternatives left over (symbolized by 'H_3')

UNIVERSITY OF WASHINGTON PRESENTATION Thursday, 21 July, 10:45-12:00 a.m.

SWARM

Dr. Swarm opened his presentation by announcing that the main goal of this project is to develop some thoughts and schemes for evaluating ranslation, rather than translation itself. He continued with a brief discussion of the following:

(1) the 650 Lexicon Format

(2) the 650 Tag Form

(3) Format for 13,000 Form Lexicon for *IBM* 709

Dr. Swarm then presented his handout, Kernel Analysis in Translation, and Translation Evaluation. He indicated that 2500 kernels had been analyzed so far. He added that the twelve most frequent Russian kernel structures account for approximately fifty percent of the occurrences.

He mentioned the fact that they are presently preparing a 13,000 Tag English dictionary, on which he spoke briefly. He then turned the presentation over to his colleague, Dr. Lytle.

LYTLE

Dr. Lytle proceeded to discuss the project's current work, insofar as the problems of multiple meaning are concerned. The human translator resolves most multiple meaning problems by looking at the context, and it would be desirable to achieve the same process by mechanical means.

UNIVERSITY OF WASHINGTON PRESENTATION

Dr. Lytle concluded the presentation with some observations concerning the solution to semantic problems by means of coordinated and dimension.

UNIVERSITY OF TEXAS PRESENTATION Thursday, 21 July, 10:45-12:00 a.m.

PENDERGRAFT

Mr. Pendergraft opened his presentation with a description of an IBM 709 computer system being programmed by the project. He explained that the system has three purposes: (a) to display generalized translation processes so that they may be tested and evaluated, (b) to assist linguists in compiling formational and interlingual data for languages to be translated by these processes, and (c) to suggest means of optimizing these processes and their data for practical applications. He elaborated on the translation process to be studied initially in the system, answering occasional questions from the group. Some of the points which received emphasis were:

(1) A generalized translation process is a process which satisfies the translation requirements of a general theory of linguistic structure, rather than merely the translation requirements of a certain pair of languages. (2)The translation process being programmed is for phrase structure languages. (3)It contains three subprocesses: recognition, transfer, and production. (4) Because recognition and production are essentially inverse processes, formation data (phrase structure grammars) for the two processes may be reorganized automatically to interchange input and output languages. (5) Any pair of languages in the system may be translated through common interlingual data. (6) The process assumes an unbroken sequence of input text and then does its own "chunking" as an integral part of recognition.

Mr. Pendergraft spent the last portion of his presentation offering graphic examples of two basic phrase structure recognition processes.

UNIVERSITY OF INDIANA PRESENTATION Thursday, 21 July, 10:45-12:00 a.m.

ELLSON

Dr. Ellson opened his presentation with a brief account of the results so far obtained by the Indiana project. He said that their basic approach was semantic but this did not mean that syntax was being (or could be) neglected. He added that a problem in semantic analysis has been that of avoiding the necessity for human coding.

He reported progress in assembling a representative sample of scientific writing. Present plans are to provide a bibliography of 25 and reproductions of 5 articles randomly sampled from abstracts published in 1959 in each of 9 scientific fields.

In the remaining time Dr. Ellson sketched an alternative to the general approaches to MT represented by the work of other projects reported at this meeting. In all of these the program for translation is basically deductive, accomplished by applying rules of dictionary equivalence, grammar, syntax, style, etc. to the source language. As evidence for the existence of an alternative, Dr. Ellson pointed to the fact of translation by people, especially children, who do not know these rules. As one alternative, he outlined in inductive approach that utilizes multiple conditional probabilities in a form of pattern analysis and gives rise to a computer or computer program which "learns" to translate by experience with the source and target languages rather than by being programmed in terms of linguistic rules.

OPEN DISCUSSION SESSION Thursday, 21 July, 2:00-5:00 p.m.

General Topic: Discussion of Possibilities of Coordinating Formats for Future Work

Chairman: Sydney Lamb

The session opened with the keynoting of the necessity for establishing a program of profitable exchange of research data between the different projects, where mutual interests can be said to exist. Professor Lamb then invited the conference participants to express their views, at the same time indicating what contributions they might be prepared to make.

The first response came from Mr. Pendergraft of the University of Texas. He offered to make their 709 programs available, including his system for the recognition process. He was of the opinion that any format should include what anyone among the represented projects should want.

Dr. Yngve once again offered to place information concerning COMIT at the disposal of all participants to whom it would be of use. Professor Lamb asked about the ways in which COMIT might be used for MT. Dr. Yngve specifically named three:

(1) for one-shot programs

(2) for developing grammars

(3) for building dictionaries

Professor Lamb added that he and his group had already written a search program in COMIT.

Dr. Yngve next offered some information about LISP (List Processor). He explained that it was a programming language useful for information processing, and that it was available in both IBM 704 and 709. He further explained that the program consisted of sub-routines for manipulating data blocks.

Dr. Yngve said that Professor John McCarthy, Computation Center, Massachusetts Institute of Technology, should be contacted for further information. He explained that Professor McCarthy had a programming manual, and that conference participants can correspond directly with him.

Professor Lamb asked Dr. Harper whether he might comment on RAND's progress with semantic coding. Dr. Harper discussed what he thought the relationship between semantic and syntactic coding should be. Concerning diagrammed text, Dr. Harper announced that RAND presently has 270,000 words.

An active discussion followed regarding the present nature of cooperation and coordination among several projects. Dr. Lieberman noted that the groups were not only having difficulty in coming together, but there indeed seemed to be a kind of repelling force. Dr. Howerton commented that on the level of intellectual exchange, there is some activity. He felt, however, that an actual exchange of material can only consist realistically of items such as word lists. Considerable controversy ensued, concerning such topics as the need for having a glossary of terms in the field of MT.

After lively debate, in which practically all participants expressed their views, Dr. See made the following motion:

> "On a voluntary basis, committees shall be formed, one for each language presently being studied, by two or more groups. One member will represent each participating project. Projects may participate whether or not they are at present working on the language in question. Said committees will meet to discuss ways, means, and practicability of agreement on format and exchange of dictionary-type information."

OPEN DISCUSSION SESSION Chairman: Sydney Lamb

Professor Lamb put the motion to a vote, and it was carried unanimously. The following are groups volunteering:

Russian: NBS, Berkeley, Wayne State U., U. of Washington, Milano and Georgetown.

French: Georgetown

The session adjourned with the proposal that the newly formed committee on Russian convene Thursday evening at 8:15.

INFORMAL SESSION CALLED BY MISS MASTERMAN Thursday, 21 July, 9:00 p.m.

Miss Masterman compared and contrasted the CLRU Mark II schema for semantic analysis, (the Thesaurus, T, a specification of which she had brought to the Conference) with the system of semantic classification proposed by Professor Ceccato, and described in a forthcoming report, a copy of which he had brought to the Conference.

She declared her reaction to the present status of M.T. research, as well as her opinion that the basic problems facing M.T. were semantic. She further noted that the only group whose work most closely paralleled that of CLRU, was Professor Ceccato and his staff. She felt that both their groups had common interests insofar as semantic analysis was concerned, even though their methods of analysis were different. In detail, she queried the advisability of using a large number of semantic classifiers.

Professor Ceccato replied to this criticism.

In open discussion, it became clear that the centrality of semantic problems to M.T. was now being widely appreciated.

OPEN DISCUSSION SESSION Friday, 22 July, 9:00-12:00 a.m.

General Topic: Discussion of Possibilities of Coordinating Formats for Future Work

Chairman: Victor H. Yngve

Dr. Yngve opened the session suggesting that it would be both profitable and of mutual interest if, during this session, time be spent making a formal listing, to include the status of existing dictionaries (including grammar coding) and analyzed text, among the various MT projects whether or not they are represented at the Conference.

As the individual projects volunteered the information, Dr. Yngve constructed two charts on the blackboard. Reproductions of these charts appear on the following two pages.

The entire session was devoted to gathering this information and to discussion of possibilities for exchanging the material.

	PROJECT	Dictionary	Text (K)	(Kind)***	Grammar Code	Govt Code	See Note*	English Equivalent
Represented at this Conference	GEORGETOWN	Russian	10,800	Stems	X	X	IP EI	1 or 2
	BERKELEY	Russian	600,000** (Biol. & Chem.)	Forms	X	X		0 - 3
	RAND	Russian	24,000 (Physics)	Forms	X		SAN IP EI	1 plus
	MILANO	Russian	600 14,000	Stems Forms		X	NS	1 - 3
	UNIV OF WASHINGTON	Russian	13,000 170,000	Forms Untagged Forms	X	X	FS	1 plus 1 plus
Not Represented	HARVARD	Russian	15,000 150,000	Stems Forms	X	X		All
	IBM	Russian	55,000 5,000	Stems Idioms	X	X		1 or 2
	RAMO-WOOLDRIDGE	Russian	15,000	Forms	X	X	IP EI	1 plus

PROJECT	Transliteration	Degree of Non-Cyrillic Punching	Text	Analysis
GEORGETOWN	1 to 1	Max plus (*)	480K (Russian) 250K (French)	-- --
BERKELEY	31 to 32	Max	30K 120K (Russian)	X --
RAND	1 to 1	Some	270K 390K (Russian)	X
MILANO	---	---	---	---
U. of WASHINGTON	---	---	---	---
WAYNE	1 to 1	Some plus	10K (Russian-English)	--
M.I.T.	1 to 1	Max	200K (English) 100K (German) 20K (German-English)	-- -- --
TEXAS	1 to 1	Min	500K (German-English)	--
C L R U	1 to 1	Min	3K (Latin)	--
(GEORGETOWN)	(C a b l e C o d e)		20K (w/concordances: Chinese, English, French, German)	
HARVARD	1 to 1	Min	25K (Russian)	--
I B M	1 to 1	Max plus	100K plus (French)	some?
RAMO-WOOLDRIDGE	1 to 1	Some	62K (Russian)	--

HANDOUTS

The following is a list of printed information and publications which were discussed in connection with the various presentations at the Princeton Conference, and which were generally distributed as handouts among the Conference participants.

GEORGETOWN

A Symbolic Lanquage for Programming the Simulated Linguistic Computer, A.F.R. Brown, Georgetown University, Machine Translation Research, MT Work Paper - Series B, No. 5.

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

Status of the COMIT System. Victor H. Yngve, Massachusetts Institute of Technology.

A Model and an Hypothesis for Language Structure, Victor H. Yngve, Research Laboratory of Electronics and Department of Modern Languages, Massachusetts Institute of Technology, Cambridge, Massachusetts, (This paper is being published in the Proceedings of the American Philosophical Society, Vol. 104, No. 5).

High-Speed Searching of Texts. K.C. Knowlton, Massachusetts Institute of Technology.

A sample copy of some output material.

BERKELEY

Diagram of the Structure of a Translation System, Sydney M. Lamb.

Types of Relationships Between Levels, Sydney M. Lamb.

Examples of Metataxemes in Representations of Lexemes, Sydney M. Lamb.

Examples of the Participation of Items of Less-Than-Word Length in Nonce Forms. Sydney M. Lamb.

Derivational Suffixes Now Being Segmented, Sydney H. Lamb.

Rough Estimates on the Quantity of Words Formable Through Combination of Segmented Lexemes, Sydney M. Lamb.

Speed and Cost of CALDIC (University of California Dictionary) System on Various General-Purpose Computers and Comparison with Brand X, *Sydney* M. Lamb.

The Form of Two Short Sentences on Each of Five Levels of Linguistic Structure, Sydney M. Lamb.

A List of Derivational Suffixes Considered for Segmentation, C. Douglas Johnson.

CAMBRIDGE LANGUAGE RESEARCH UNIT

What Is A Thesaurus?, Margaret Masterman, June 1959, ML90 i.

A Flexible Punched-Card Procedure for Word Decomposition, M. Kay and R. Mackinnon Wood, M.L. 119.

Comments by A.F. Parker-Rhodes on "Transformations & Discourse Analysis Projects", No. 12, University of Pennsylvania Workpaper on the Syntactic Analysis of English, March 1960.

Information Retrieval Term List. CLRU, ML/131.

Mechanical Pidgin Translation. Margaret Masterman and Martin Kay, ML/133.

A Commentary on the RAND, Sentence Structure Determination Program, A.F. Parker-Rhodes, ML/134.

The Information Retrieval System of the CLRU, R.M. Needham, A.H.J. Miller, K. Sparck Jones, ML/109.

Notes on Making Dictionary Entries for the CLRU Bracketing Program, A.F. Parker-Rhodes.

First Sample of Dictionary Entries for the CLRU Bracketing Program, A.F. Parker-Rhodes.

WAYNE STATE UNIVERSITY

Research in Machine Translation, Russian to English, Wayne State University.

RAND CORP

Available RAND Linguistic Data

Analytic Reports. Project 4116/2905, MT 1-14-60:CHS.

NATIONAL BUREAU OF STANDARDS

Recognition of Clauses and Phrases in Machine Translation of Languages, Franz Alt, NBS Report 6895.

The Outlook for Machine Translation, Franz Alt, National Bureau of Standards, (From "Proceedings of the Western Joint Computer Conference", Vol. 17, San Francisco, California, May 1960.)