MECHANICAL TRANSLATION
DEVOTED TO THE TRANSLATION OF LANGUAGES WITH THE AID OF MACHINES

VOLUME ONE, NUMBER THREE DECEMBER, NINETEEN FIFTY-FOUR


## NEWS

### SUBSCRIPTIONS

It has become necessary to place MECHANICAL TRANSLATION on a subscription basis. The subscription price for Volume Two (1955) will be $1.00, which will cover about half the estimated cost of printing and distribution for three numbers. We are looking for support for the balance.

Volume One was distributed free of charge as an experiment to see how great the interest would be, and the cost was borne by the Massachusetts Institute of Technology. The mailing list now numbers about 950.

One measure of the success of the experiment will be the number of subscriptions that come in. If you wish to receive Volume Two for 1955, please fill in the enclosed form and send, together with check or money order for $1.00 (payable to W. A. Hokanson, Bursar, M.I.T.), to:

MECHANICAL TRANSLATION
Room 14N-307
Massachusetts Institute of Technology
Cambridge 39, Massachusetts

### NEW COURSE

Anthony Oettinger and Lawrence G. Jones are offering a seminar in Mathematical Linguistics during the spring term at Harvard. The subject of discussion will be the application of mathematical tools to the study of the structure of spoken and written language, and to the analysis of the problem of meaning. Recent developments in the automatic translation of languages, in speech analysis and synthesis, and in theories of communication will be considered in detail.

### BOOK ON MT

Scheduled for spring publication jointly by the Technology Press of M.I.T. and John Wiley & Sons, Inc..MACHINE TRANSLATION OF LANGUAGES is a collection of up-to-date essays edited by W. N. Locke and A. D. Booth and written by seventeen workers in the field. The book has a preface by Warren Weaver. It also contains his original memorandum "Translation" and a historical introduction by A. D. Booth and W. N. Locke.

### STATISTICAL LINGUISTICS

A 123-page bibliography on statistical linguistics has recently been published. It is: BIBLIOGRAPHIE CRITIQUE DE LA STATISTIQUE LINGUISTIQUE by Pierre Guiraud, revised and completed by Thomas D. Houchin, Jaan Puhvel and Calvert W. Watkins of Harvard University under the direction of Joshua Whatmough, Editions Spectrum, Utrecht/Anvers, 1954.

### ADAPTIVE SYSTEMS

A preliminary informal bibliography on adaptive systems and automata has been prepared by F. A. Webster, and may be obtained by writing him at 62 Coolidge Avenue, Cambridge, Mass. Besides general articles, the bibliography includes learning models, game-playing devices, pattern recognizers, and physiological and psychological studies.

### EIGHT ARTICLES ON TRANSLATION

The INTERNATIONAL JOURNAL OF AMERICAN LINGUISTICS has devoted its entire fourth

Continued on page 55

# THE DISTRIBUTION OF WORD LENGTH IN TECHNICAL RUSSIAN

Anthony G. Oettinger
Computation Laboratory, Harvard University

IN the course of an analysis of several samples of technical Russian undertaken as part of a study in mechanical translation, a number of statistical data reflecting the structure of these samples were compiled. One of these, the distribution of word length, is presented here as Fig. 1.

The theoretical interest of this distribution arises from the possibility of using it as a basis for an operational definition of words in printed texts. If texts are considered purely as sequences of symbols including the letters, punctuation marks, and space, the resulting sequences are of a length which no practicable machine can manage. A study of the distribution of the number of symbols between pairs of successive symbols of certain classes would be one way to reveal structural characteristics of the text sequences potentially useful toward the definition of manageable and significant subsequences. The subsequences included between successive occurrences of letter pairs have not been investigated. Those included between successive pairs of periods, exclamation points or question marks can be identified with the classical sentence, and finally, those included between successive pairs of punctuation marks or spaces can be identified with words. The length distribution of the latter subsequences has the desirable property, not shared by the others, of being concentrated at relatively low values of length, and of having no elements exceeding a certain length (Fig. 1). Words, defined in this fashion, can readily be identified by a machine and they are of limited variety, so that their listing in a dictionary is practicable.

From the practical point of view, the distribution is useful in planning input and storage facilities in experimental translating equipment.

The samples used were relatively small, and Fig. 1 should therefore be interpreted with great caution. The bar graph represents the distribution of a sample totalling 6,486 words. Points are used to indicate the distributions obtained from smaller constituents of the total. The scattering is such as to indicate that samples 1, 2, and 3 differ significantly among each other in details of their distributions. An examination of the texts indicates that these differences can safely be attributed to differing subject matter and styles. However, all distributions are bimodal, perhaps trimodal, and cut off at k=18. The mode about k= 7 is attributable to the large number of different words used to define the particular subject of each text. The peaks at k= 1 and at k= 3 are due to a small number of very frequent "grammatical words," that is, prepositions, conjunctions, etc. The five most frequent words of length 1, 2, and 3 in the total sample are listed in Table 1. This table shows that the most frequent two letter words are consistently less frequent than three letter words of similar rank. One and two letter words are exclusively grammatical; 90% of the three letter words are also grammatical, leaving 10% dependent on the subject matter. The words of length 4 are nearly all inflected. The fact that only very few Russian words have stems of three or less letters probably accounts for the valley at k= 4. Indications thus are that the modal and cut-off structure of the distributions are functions of the structure of the Russian language, while variations within these structures are characteristic of individual authors. For those who might wish to draw their own conclusions, the raw data is given in Table 2, and the sources of the samples are listed in Table 3. Letter, diagram and suffix distributions compiled from the same samples may be found in the reference.

TABLE 1

| | | | | | |
|---|---|---|---|---|---|
| v | 210 | na | 86 | pri | 93 |
| i | 165 | iz | 57 | dlja | 72 |
| s | 91 | po | 46 | chto | 50 |
| k | 43 | ot | 28 | kak | 29 |
| a | 21 | ne | 26 | ili | 22 |

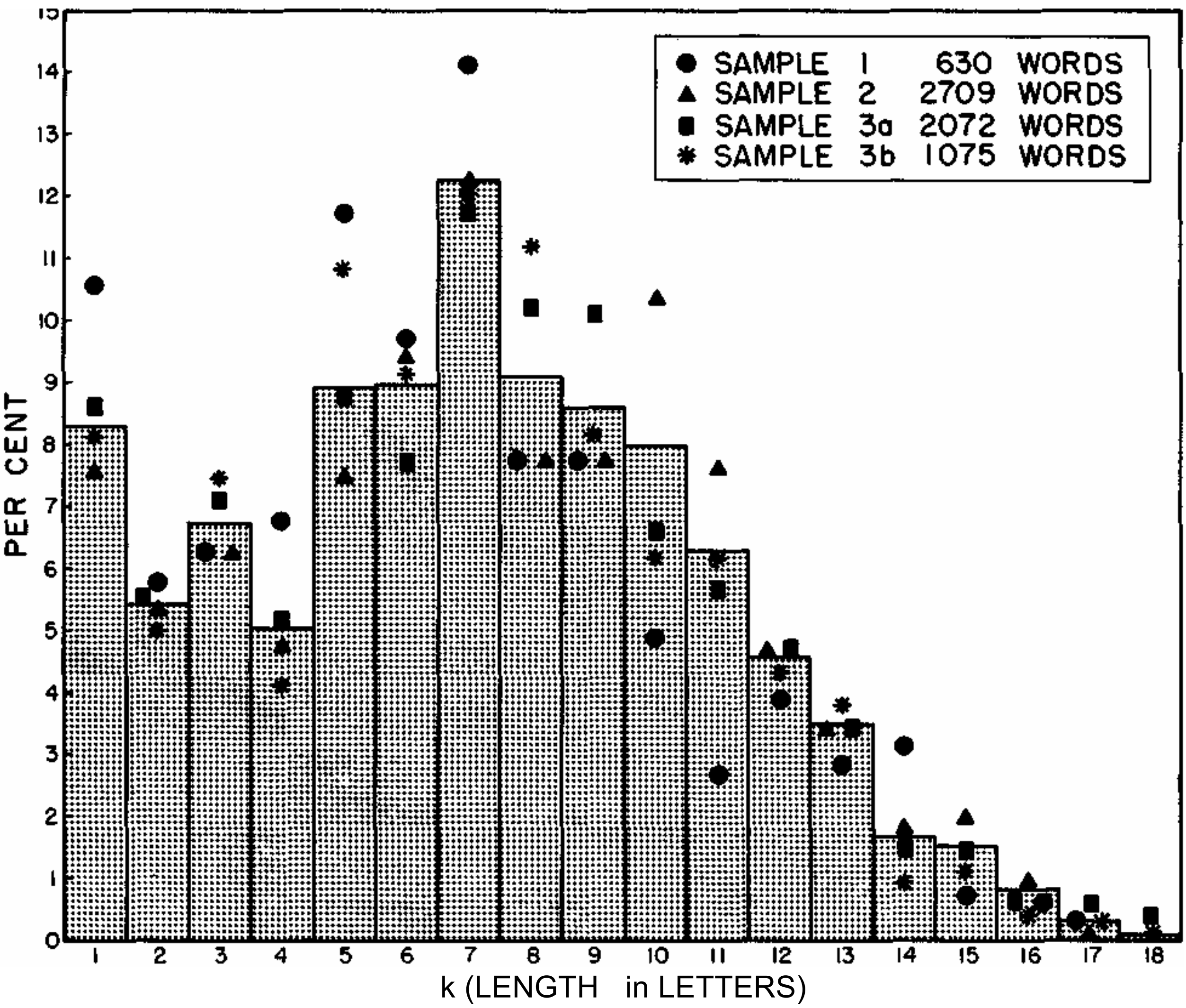
SAMPLE 1 630 WORDS
SAMPLE 2 2709 WORDS
SAMPLE 3a 2072 WORDS
SAMPLE 3b 1075 WORDS
PER CENT
0
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
k (LENGTH in LETTERS)

Figure 1

TABLE 2

| Word length | Frequency | | | | |
|---|---|---|---|---|---|
| | Sample 1 | Sample 2 | Sample 3a | Sample 3b | Total |
| 1 | 67 | 204 | 178 | 88 | 537 |
| 2 | 36 | 147 | 114 | 54 | 351 |
| 3 | 40 | 170 | 148 | 80 | 438 |
| 4 | 43 | 130 | 107 | 45 | 325 |
| 5 | 74 | 203 | 183 | 117 | 577 |
| 6 | 61 | 258 | 161 | 99 | 579 |
| 7 | 89 | 332 | 245 | 129 | 795 |
| 8 | 49 | 209 | 212 | 121 | 591 |
| 9 | 49 | 209 | 211 | 88 | 557 |
| 10 | 31 | 281 | 138 | 67 | 517 |
| 11 | 17 | 208 | 118 | 66 | 409 |
| 12 | 25 | 127 | 98 | 47 | 297 |
| 13 | 18 | 94 | 72 | 41 | 225 |
| 14 | 20 | 50 | 29 | 10 | 109 |
| 15 | 5 | 54 | 28 | 13 | 100 |
| 16 | 4 | 28 | 16 | 5 | 53 |
| 17 | 2 | 5 | 9 | 4 | 20 |
| 18 | 0 | 0 | 5 | 1 | 6 |

TABLE 3

1. A. G Lunts, 1950, "Prilozhenie Matrichnoj Bulevskoj Algebry k Analizu i Sintezu Relejno-Kontaktnyx Sxem," Doklady Akademii Nauk SSSR, 70, pp. 421-23.

2. K. V. Valdimirskij, 1951, "O Sinxronnom Fil'tre," Zhurnal Eksperimental'noj i Teoreticheskoj Fiziki, *21,* pp. 2-10.

3. B. P. Aseev, 1947, Osnovy Padiotexniki (Moskva: Svjaz'izdat) (a) pp. 10, 18, 20, 21, 23, 33, 37, 42, 45, 49, 55 (part); (b) pp. 55 (part), 59, 64, 65, 71, 122

## REFERENCE

Oettinger, A. G., "A Study for the Design of an Automatic Dictionary," Doctoral Thesis, Harvard University (1954).

# COMPUTING MACHINES FOR LANGUAGE TRANSLATION

T. M. Stout*
Schlumberger Instrument Company
Old Quarry Road, Ridgefield, Conn.

RESEARCH on the problems of machine translation has been going on for several years in this country and abroad.[1] To date it has been concerned primarily with the complicated linguistic problems involved in mechanical translation, since the engineers can probably build the necessary equipment. This article is intended to suggest some of the linguistic problems to the engineer and to explain some of the engineering ideas for the amateur or professional linguist. The reader is cautioned that the procedures and equipment described are not necessarily the best or most recent, and that considerable development must be done before an actual mechanical translator is built and put into operation.

## General Approach: The Language Problem

Present proposals for a mechanical translator involve, in rough terms, constructing a machine which carries out automatically the process that the human translator is imagined to use in converting a sentence from one language (the <u>input</u> language) into a new language (the <u>output</u> language). This process is assumed to consist of (1) transferring the material from the printed page to the brain (reading); (2) searching a dictionary to establish the meaning or meanings of each word in the original text; (3) selecting the correct meaning from the possible alternatives; (4) rearranging and refining the results to fit the requirements of the output language; and (5) recording of the results in written or other form for future use. The general procedure may be illustrated by an example.

Suppose the translator is faced with the German sentence:

<u>Er fand die Aufgabe zu schwer,</u>

which may be translated, "He found the task too difficult." A German-English dictionary gives the following meanings for the individual words:

<u>Er</u> - he
<u>fand</u> (from <u>finden</u>) - found;thought,considered
<u>die</u> - the (article); that, this, he, she, it (dem. pronoun); who, which, that (rel. pronoun)
<u>Aufgabe</u> - task, duty; lesson, exercise; asking (of riddle); posting (of letter); registration (of luggage); giving up, shutting down (of business)
<u>zu</u> - to, at, in, on (preposition); too (adverb)
<u>schwer</u> - heavy; oppressive; clumsy; difficult; grave (illness); indigestible (food); strong (cigar)

<u>Er</u> can be translated only by "he." Although <u>finden</u> generally means "to find" in the sense of "to discover," it also has the figurative meaning, "to think" or "to consider." English "find" also shares these meanings and no great harm will be done if <u>finden</u> is always translated as "find." The presence of a noun following <u>die</u>, indicated by the capital letter or by a dictionary entry opposite <u>Aufgabe</u>, makes its translation "the." The translation of <u>Aufgabe</u> may be taken as "task" in all cases, since this meaning is general enough to include all of the other, specialized meanings; the nature of the task should be clear from the context. <u>Zu</u> is translated as "too" because of the following adjective, which presents the toughest problem in the sentence. The choice in this case evidently depends on the feeling that a task can be difficult, but not heavy, clumsy, grave, indigestible, or strong.

As this meaning suggests, a word which has only one meaning (or can arbitrarily be assigned only one meaning) will present no problems. Any

* This work was done at the Department of Electrical Engineering of the University of Washington in Seattle, Washington, and was originally published in THE TREND in Engineering at the University of Washington, Vol. 6, No. 3, p. 11 ff, July 1954. The author's interest in mechanical translation and many of the ideas contained in this article are the result of conversations with Dr. Erwin Reifler of the Far Eastern Department of the Univer-- sity of Washington.

[1] MECHANICAL TRANSLATION, Vol. I .March 1954, published at the Massachusetts Institute of Technology. An extensive bibliography of publications in this field.

word with several meanings, however, will cause considerable trouble. The selection of a particular meaning is sometimes based on grammatical considerations, sometimes on the presence of other words or types of words, and sometimes on the nature of the subject matter. In addition to the ability to read and write and search a dictionary, the machine - like the human translator - must be able to discern gramatical distinctions and the occurrence of words which determine the meanings of associated words.

## Coding

At the present stage of development, it is assumed that the translating machine will work only with printed material. In addition to some obvious engineering advantages, this approach has the linguistic advantage that the written language is more distinctive than the spoken language. In English, for instance, the homonyms, not-knot, pair-pear-pare, and numerous other groups of words are easily distinguished by their spelling. The number of words with the same spelling and different pronunciations,such as lead-lead and bow-bow, is much smaller.

Since most computers are designed to work with numbers, the incoming text must be con verted from the written alphabet into a numerical form acceptable to the machine. Several different coding schemes are available for this purpose. One obvious procedure is simply to number the letters, using either two-digit deci mal numbers or five-digit binary numbers. Coded in this manner, A-B-C-D. . .would be come 01-02-03-04..., or 00001-00010-00011-00100…

Other codes are commonly used in standard equipment which might be incorporated in a translating machine. Machines available from IBM use the code given in Table I, in which each letter is represented by two holes punched in a column of a standard punched card; the upper hole is called a zone punch and the other is a digit punch. Standard teletypewriters use the Baudot code given in Table II, which employs five pulse positions in a manner similar to the binary code (plus a sixth pulse for timing).

Binary or teletype coding requires more digits for each letter than the decimal or IBM coding and might appear to require considerably more space. On the other hand, these codes employ only two symbols (0 and 1, pulse and no pulse) for each digit. The physical elements in the computer can therefore be simple two-state devices, such as a switch or relay whose contacts are either closed or not closed, a vacuum tube which does or does not carry current, a magnetic core which is magnetized or not, and so forth. Since it is easy to determine which state exists, reliable operation is obtained without any accurate measurements or precision components.

**TABLE I**

**ALPHABET CODING USED IN IBM PUNCHED CARD EQUIPMENT**

| | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S | T | U | V | W | X | Y | Z |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 11 | x | x | x | x | x | x | x | x | x | | | | | | | | | | | | | | | | | |
| 12 | | | | | | | | | | x | x | x | x | x | x | x | x | x | | | | | | | | |
| 0 | | | | | | | | | | | | | | | | | | | x | x | x | x | x | x | x | x |
| 1 | x | | | | | | | | | x | | | | | | | | | | | | | | | | |
| 2 | | x | | | | | | | | | x | | | | | | | | x | | | | | | | |
| 3 | | | x | | | | | | | | | x | | | | | | | | x | | | | | | |
| 4 | | | | x | | | | | | | | | x | | | | | | | | x | | | | | |
| 5 | | | | | x | | | | | | | | | x | | | | | | | | x | | | | |
| 6 | | | | | | x | | | | | | | | | x | | | | | | | | x | | | |
| 7 | | | | | | | x | | | | | | | | | x | | | | | | | | x | | |
| 8 | | | | | | | | x | | | | | | | | | x | | | | | | | | x | |
| 9 | | | | | | | | | x | | | | | | | | | x | | | | | | | | x |

**TABLE II**

**STANDARD BAUDOT TELETYPE CODE**

| LETTER | PULSE 1 | 2 | 3 | 4 | 5 | LETTER | PULSE 1 | 2 | 3 | 4 | 5 |
|---|---|---|---|---|---|---|---|---|---|---|---|
| A | X | X | | | | N | | | X | X | |
| B | X | | | X | X | O | | | | X | X |
| C | | X | X | X | | P | | X | X | | X |
| D | X | | | X | | Q | X | X | X | | X |
| E | X | | | | | R | | X | | X | |
| F | X | | X | X | | S | X | | X | | |
| G | | X | | X | X | T | | | | | X |
| H | | | X | | X | U | X | X | X | | |
| I | | X | X | | | V | | X | X | X | X |
| J | X | X | | X | | W | X | X | | | X |
| K | X | X | X | X | | X | X | | X | X | X |
| L | | X | | | X | Y | X | | X | | X |
| M | | | X | X | X | Z | X | | | | X |

## Input and Output Devices

A number of standard devices are available for coding the incoming text for insertion into the machine and, after the translation process is completed, for decoding and printing the translation in the output language. Teletypewriters, operated by typists with no knowledge of either language, could be used to supply electrical signals directly to the translating machine or to prepare punched paper tape for later use. Similar machines can be used to type the final output of the translator.

Input devices now available are relatively slow, so that faster means of supplying material to the translating machine would be essential. An electronic reading device, capable of working directly from the original printed text, has recently been announced.[2] Faster output devices will also be required to maintain over-all balance.

## Storage

The dictionary needed in a mechanical translating machine might be stored on a magnetic drum such as the one shown in Fig. 1. This type of storage, in which information is stored by magnetizing small areas on the surface of a revolving cylinder, is widely used in arithmetic computers and has a number of desirable properties: a large ratio of information to volume, lower access time, permanence, and simplicity.

Individual words are stored along the length of the drum (each letter being represented by a group of five magnetized or unmagnetized spots) and pass the reading heads once in each revolution of the drum. Words in the input language are stored at one end of the drum, and their equivalents in the output language at the other end. If the drum is rotated at 2,400 rpm, or 40 rps, each word is available in not more than 25 milliseconds. Following standard practice, 80 spots per inch can be placed around the circumference of the drum and 8 tracks per inch along the length of the drum. Allowing 10 letters or 50 tracks per word in both halves of the dictionary, a drum 12.5 inches long and 12 inches in diameter would hold approximately 3,000 words and their translations.

In order to reduce the average time spent in searching the dictionary, certain common words might be stored several times on the same drum. The 850-word vocabulary of Basic English could be stored three times on a single drum, so that any particular word is available

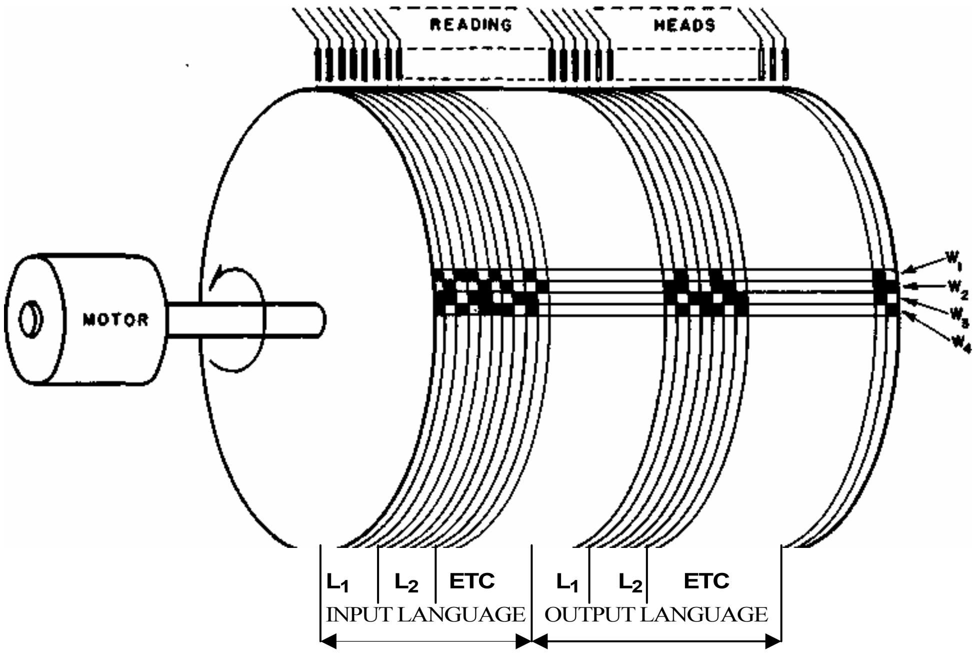


FIG. 1. MAGNETIC DRUM FOR DICTIONARY STORAGE

Words ($W_1$, $W_2$, etc.) are stored along the length of the drum,and each letter ($L_1$, $L_2$, etc.) requires five tracks around the drum.

2 Shepard, D. H., "The Analyzing Reader." A paper presented at the IRE convention in San Francisco, Aug. 19, 1953.

in a third of a revolution or less (not over 8 milliseconds).

To provide an adequate vocabulary for satisfactory translation, several such drums would be required. By searching all drums simultaneously, as explained below, any word in the dictionary could be found in the time required for one drum revolution. At approximately one cubic foot per drum, exclusive of the associated circuits, the space required for a vocabulary of 100,000 words or so becomes rather large. A number of tricks are available, however, for reducing the size of the mechanical dictionary.

If we are concerned with translation into English, as seems probable, many words in the input language text will not require translation. English has borrowed extensively from other languages and many foreign words are immediately recognizable by the English reader. A glance at a German dictionary, for example, reveals such words as Deck, Despot, Diplomat, and Dock which are identical with the English forms; we also find Demagog, Demokrat, direkt, Distanz. and Doktor which differ slightly in spelling but would present no real difficulties to the reader. The translation process can be by-passed for such words, and the original input word printed directly in the output. This approach must be used with caution, since the two languages may not share all the meanings and connotations of a given word, but it does offer hope for tremendously reducing the size of the mechanical dictionary.

Compound words are rather common in German and can, in fact, be invented at will by writers and speakers. If the meaning of a compound is clear from the meanings of its constituents (as is likely for all except old well-established compounds, which will be entered as distinct words), the dictionary can be searched for each constituent separately, and the respective translations compounded on the output side.

Endings, used extensively in other languages to convey grammatical information such as tense and number, can be treated in similar fashion to effect a further reduction in the size of the dictionary. Each word might be regarded as a compound built from a stem, common to all forms of the particular word, and an ending, which may be shared with other words. The dictionary may then be split into a large stem section and a small ending section. A useful by-product of this procedure is the grammatical information made available by the identification of an ending; this may be used in the elimination of impossible translations, discussed hereafter.

The techniques used in the dissection of compounds will be valuable in still another way. If a word has more letters than are permitted by the physical size of the dictionary (ten letters in the example above), it can be split into two parts which separately signify nothing. Beratschlagen. for example, might be split into Beratsc and hlagen, with parts of the translation stored opposite each half. Dictionary space is used more efficiently in this manner, but the processing time may be increased excessively.

Splitting words in order to determine parts of a compound, or stems and endings, is fraught with difficulties which must be explored by linguists. The engineering techniques for carrying out these operations have been devised, but are too involved to discuss here.

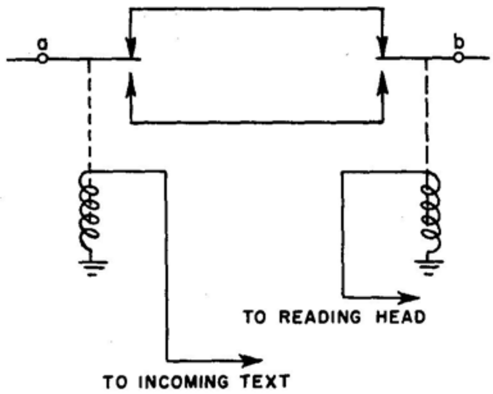


FIG. 2. RELAY CIRCUIT FOR CHECKING ONE PULSE OF INCOMING TEXT WITH ONE TRACK ON DICTIONARY

## Dictionary Search

In making a mechanical translation, the first step is a comparison of each word of the incoming text with the entire dictionary. If any word is not found in the dictionary in its original form, the dissection scheme for endings and compounds can be tried; if this fails, the word can be printed through without alteration.

Several methods are available for making this comparison; an impractical but easily understood system is shown in Fig. 2. This system requires two single-pole double-throw relays for each pulse position: one relay operated by the incoming text and the other relay operated by pulses from the reading heads on the magne-

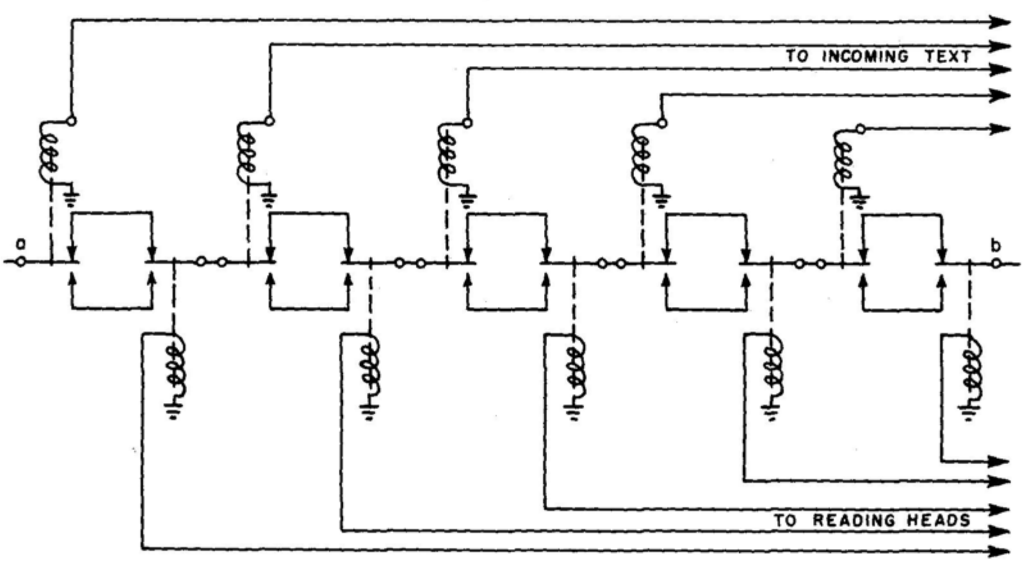


FIG. 3. COMBINATION OF RELAYS FOR CHECKING ONE LETTER

tic drum. The path between points "a" and "b" is closed only when both relays are either energized (pulses present in both incoming word and dictionary) or not energized (spaces present in both places). The occurrence of a closed path, therefore, indicates that the particular pulse position is identical in both the incoming word and the dictionary.

Entire letters, coded as a group of five pulses or spaces, can be checked by a series combination of five such relay circuits, as shown in Fig. 3. In corresponding fashion, words of ten letters could be checked by a series combination of fifty such relay circuits. A closed path through a long string of such circuits indicates that the incoming word has been found in the dictionary, and this event can be made to initiate printing of the translation stored at the other end of the drum.

An input-language word with several meanings can be entered in the dictionary several times, each time with a suitable translation. The searching procedure outlined above would uncover each of the possible translations and would make them all available for further consideration. To assist in the subsequent selection of one of these meanings, each translation might have a "tag" stored with it, which would supply grammatical or other necessary information needed by the machine.

With a multiplicity of such circuits, a number of dictionary drums could be searched simultaneously, as suggested schematically in Fig. 4. The incoming text is supplied to all drums at the same time. Correspondence between the incoming word and a dictionary entry is noted on only one drum, from which the translation is obtained. Parallel operation of this type would permit a dictionary of any desired size with the access time of a single drum, but at a considerable price in additional checking circuits.

In a practical comparison system crystal diodes, transistors, or vacuum tubes would be used instead of relays. These elements have no moving parts to limit the speed of operation and require much less signal power.

## Multiple Meaning

Having obtained the possible translations for each word in a sentence, the machine is faced with the problem of selecting the correct meaning from several alternatives. This problem can be attacked in a number of ways.

In technical writing many words have specialized meanings which are used in all texts in a given area of science. For example, Flügel in a paper on aeronautical engineering is much more likely to mean "wing" than "grand piano," both of which are given in a general dictionary. The machine could be instructed to select the specialized meaning when the text is known to be in a specialized area (by means of appropriate tags) or special dictionaries could be used.

A number of distinct problems can be recognized in the case of general language. As indicated by the examples, the translation of a word

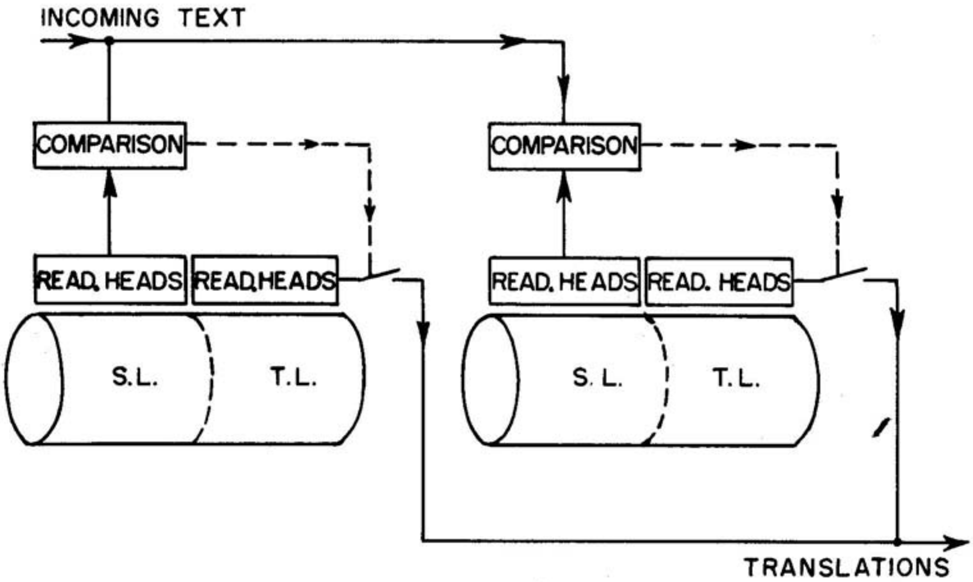


FIG. 4. SCHEMATIC DIAGRAM SHOWING SIMULTANEOUS SEARCH OF TWO DICTIONARY DRUMS

is sometimes based on grammatical considerations, sometimes on the co-occurrence of another word or type of word in the same sentence or clause, and sometimes on the larger context. In all cases, the choice is determined by examining the surrounding words and, according to rules furnished by the linguists, either selecting or eliminating certain alternatives.

The general procedure employed by the machine in selecting the proper meaning can be indicated by an example. For the German sentence given above, a superficial study suggests the following rule for the translation of zu: if zu is followed by an adjective or adverb, its meaning is "to," but otherwise it is a preposition, and its meaning must be determined by additional analysis. The translating machine can be instructed to examine the tag on the word following zu and, if the code designation for an adjective or adverb appears, to select "too" as the meaning.

Not all words present difficulties with multiple meanings, and the mechanical translator can easily locate the trouble-makers in any sentence by counting the alternatives encountered in the dictionary search. Having found a word with several possible meanings, the machine can refer to a list of rules appropriate to this word or its general class of words. This list should be flexible, so that rules can be added or discarded without disrupting the operation of the other rules. The machine can probably be arranged to count the number of times each rule is used and the number of successes scored, so that the effective rules can be applied first and ineffective rules discarded.

The linguistic rules will necessarily be coded and could, in fact, be expressed in algebraic fashion by the techniques of symbolic logic.[3] The resulting algebraic expressions can be simplified by formal procedures and can be converted directly into devices which carry out the selection process. The so-called logic circuits needed in a mechanical translator are employed in conventional arithmetic computers and their design should pose no special problems.

## Conclusion

Experiments with word-by-word translation by mechanical means have already been conducted with surprisingly good results, even where no attempt has been made to deal with the problem of multiple meanings. With even a rudimentary set of rules for selecting or eliminating some of the possible meanings, still better results should be obtained. If the linguists can discover the rules, the engineers are ready to build the equipment, given the necessary support. Practical mechanical language translation is a definite possibility for the near future.

---

3 Langer, S. K., AN INTRODUCTION TO SYMBOLIC LOGIC: New York, Dover Publications, 1953.

# THE CONFERENCE ON MECHANICAL TRANSLATION*
## Held at M.I.T., June 17-20, 1952

A. C. Reynolds, Jr.
International Business Machines Corporation, Endicott, N. Y.

The following report was prepared immediately after the writer's return from the conference. It was written from the viewpoint of an engineer listening to experts in a field far separated from his own. Such judgments as may be found interspersed amongst the reports of individual papers are of an engineering nature, and are not to be construed, as being based upon other than an amateur's knowledge of linguistic theory. Further, they represent only the reporter's evaluation, not necessarily that of his company as a whole. It is of interest, however, that the writer's company, The International Business Machines Corporation, has jointly sponsored with Georgetown University a successful demonstration of syntactically correct mechanical translation from Russian into English. The computer employed was the IBM 701, and the programming techniques used were first discussed at the 1952 conference.

The concept of mechanical translation originated in two areas, the first being cryptographic work conducted by various governments during the late war, and the second being the successful inauguration and employment of the simultaneous translation schemes presently employed by the UN and other internation conferences. Broken down into basic essentials, translation consists of memory scanning for identification of meaning in two different symbolic systems, called languages, and simultaneous editing by the translator to convert the syntactical relationships of the language being translated to those of the translated language. Of these, the memory scanning is definitely paralleled in computer techniques. If one to one correlations in meaning existed between words of different languages, programming on existing computers would be completely successful. Syntactical relationships and shading of meaning by the context of the words makes the problem of mechanization exceedingly difficult in the absence of a mechanical means of converting from one syntax to another.

Much work was stimulated by a memorandum, Translation, written by Dr. Warren Weaver of the Rockefeller Foundation.which was distributed to a selected group of linguists, psychologists, computer engineers, and philosophers. Dr. Yehoshua Bar-Hillel, acting under a grant from the Rockefeller Foundation and then conducting his research at M.I.T., acted as the coordinator of the groups actively interested in mechanical translations. As part of his work, Dr. Bar-Hillel prepared a summary entitled "Present Interest in Mechanical Translation," listing the individuals actively working on the application of computers and computer techniques to mechanical translation. In 1952 he organized a Conference on Mechanical Translation at M.I.T.

This report is concerned with providing a precis of the papers and discussions at the Conference.

Session I - June 17, 1952
Public Session

The Public Session of the Conference on Mechanical Translation was announced by invitations extended by Dr. Yehoshua Bar-Hillel to persons who might be interested in the problems of mechanical translation and, in particular to members of the Conference on Speech Communication which immediately preceded the Conference on Mechanical Translation. At the public session papers were not presented, but short talks were given by each of the five participants outlining their work in the field and their tentative proposals for future work.

Dr. Bar-Hillel discussed the need and possibilities for mechanical translation, the need primarily arising in the fields of science and of diplomacy, for analysis of popular periodicals of various countries. Although a person may be versed in the cultural or popular language of several countries, this does not necessarily mean that the same individual is capable of translating scientific treatises originating in

* For a linguist's view of the same Conference, see MT, Vol. I, No. 2, "Report on the First Conference on Mechanical Translation," Erwin Reifler, pp. 23-32. A list of participants in the Conference appears on p. 24 of that article.

the same countries. This is due to the well known fact that each scientific discipline creates its own jargon, assigning very specific meanings to common words of the language, these meanings being peculiar to the particular science itself. There is, therefore, a need for translators who are capable of making meaningful interpretations, not only in the more popular writings, but also in specific areas of scientific research. The volume of material appearing in popular periodicals is appalling in its magnitude and complete scanning of a particular nation's output is virtually impossible as long as human translators must be relied upon. He concluded that it is in these areas that mechanical translation is capable of making a major contribution to society.

Prof. Leon Dostert, Director of the Institute of Languages and Linguistics, Georgetown University, Washington, D. C., spoke on the subject of human translation versus machine translation. Prof. Dostert drew on his experience in setting up the translation system employed at the Nuremburg trials in Germany and in working with IBM in the development of the simultaneous translation system used at the UN and other international conferences. In discussing this problem, he made the statement that, except in the very specialized areas discussed by Dr. Bar-Hillel, there is no shortage of human translators, owing apparently to the fact that the current workload is regulated by their availability. The contribution a machine can make is in the processing of the vast amount of material that is currently not even being touched in the specialized fields. He described systems employed in setting up efficient simultaneous translation systems and also rapid printed translations in international gatherings. These systems were remarkably similar in their organization to machine organization for computer application. He confessed that he came to the Conference as a sceptic. (Later in the Conference he became convinced that mechanical translation would be possible.)

Dr. Olaf Helmer, Director of Research, Mathematical Division, Rand Corporation, Santa Monica, California, discussed the structure of the problem of mechanical translation. Meanings of particular words and phrases may be idiomatic or may be changed or modified by the context in which they appear. Further, each group of languages has its own syntactical relationships which are peculiar to the group,and most frequently also vary in minor details among members of the same group. The machine must be capable of resolving idiomatic, contextual, and syntactic ambiguities if human editing is to be kept at a minimum and maximum intelligibility is to be achieved. Dr. Helmer discussed schemes that have been tentatively investigated by the Rand Corporation for sol-ving this problem. His conclusion is that high speed general purpose computing machines will be able to handle the main translation task.

Dr. Andrew D. Booth, Director, The Electronic Computer Section, Birkbeck College, University of London, discussed the popular misconceptions covered by the question, "How intelligent can a machine translator be ?" The conclusions necessarily were that "intelligence" as applied to machines involves a complete mis-understanding both of intelligence and of ma-chines. No intelligence is required, on the part of the machine at least, in mechanical transla-tion.

Dr. James W. Perry, Center of International Studies, M.I.T., discussed machine techniques and index searching and translation. The basis of Dr. Perry's talk was the index searching machine developed by IBM to solve the problem of scanning vast amounts of information and ex-tracting certain specific items. He discussed the development of coding on punched cards in order to employ a machine at maximum efficiency. He concluded on the basis of his acquaintanceship with existing machines and machine techniques that mechanical translation was not only feasible but far closer to realizations than possibly the audience recognized.

A period of discussion from the floor followed the presentation of the talks. There was general agreement on the part of both the panel and the audience that mechanical translation was feasible. It was interesting to note that the computer engineers present presented all of the difficulties standing in the way of producing a mechanical translator from the engineering standpoint; the linguist, from his standpoint; and the psychologists and philosophers from the standpoint of their respective disciplines. Each agreed, however, that, if the other two groups did their work, we could in the near future produce adequate and intelligible machine programmed translations.

Session II - June 18, 1952
Chairman - Dr. Leon Dostert

Prof. Erwin Reifler.Far Eastern and Russian Institute, University of Washington, Seattle, Washington, presented the first two papers of the morning session entitled, "Mechanical Translation with Pre-editing,"and "Writing for Mechanical Translation."

The first paper concerned itself with the fact that syntactical relationships differ amongst languages. For ease in programming on a mechanical translator, a source language should be arranged according to the syntax of the target language (language into which the material is being translated). Where this is not possible due to the fact that the syntax is inseparable from the actual word form (such as the dative case in Latin) certain keys, such as capital letters or diacritical marks, can be inserted as recognizable signals for a machine whose input is a print scanning device. Pre-editing then would imply the use of a human editor to rearrange the source language insofar as possible in accordance with the syntax of the target language, and secondly, employment of various inserted signals to notify the machine of syntactical arrangements inseparable from the word form.

The second paper, on "Writing for Mechanical Translation," would necessitate the training of all writers, and more particularly their secretaries,in the required conventions for arrangement of an article for translation into a given language. The discussion of these two papers indicated that the use of a pre-editor, rather than educating all authors and all secretaries in techniques of writing for mechanical translations, is far preferable. As a matter of fact, a person skilled in keyboard operation could be readily trained to insert syntactical recognition signals at the time of keying the text into the machine. This, of course, also holds for the preparation of a manuscript for machine scanning.

Dr. Yehoshua Bar-Hillel presented a paper on Mechanical Translation employing a post-editor. Since a one-to-one correlation does not exist between meanings of words expressing essentially the same idea in various languages, if a machine operates on a comparison basis only, or even if it is capable of computing syntactical relationship, a multiplicity of words in the target language can be derived for any single word of the source language. For a particular sentence, say of 10 words length, this can easily result in possible combinations of words in the target language extending to several thousands of more or less meaningful combinations. It is necessary, therefore, to incorporate some form of post-editing in order to resolve the ambiguities inherent in this relationship between languages. Dr. Bar-Hillel is much concerned with the tremendously increased demands in terms of machine storage capacity which this situation implies. It is, however, not quite so grave as appears on the surface, since particularly in scientific writings, a vast number of one-to-one correlations do exist.

(The subject of glossaries to handle the scientific translations was covered in a later session of the conference.)

The fourth paper, "Model English for Mechanical Translation" was presented by Prof. Stuart C. Dodd, Director, Washington Public Opinion Laboratory, University of Washington, Seattle. Dr. Dodd's paper concerned itself with the standardization of English syntax as a means of simplifying the use of English either as a source language or as a target language. A model language, as defined by Dr. Dodd, means any language in which the rules of syntax have been regularized, and in which familiarity of words is a governing criterion. The specific rules used in regularizing a language are itemized in the paper. The examples employed by Dr. Dodd indicate that regularizing, that is, constructing a model language, impaires but very slightly the readability and understandability of the subject matter. In English, at least, regularizing leads only to a certain quaintness of expression somewhat similar to the sentence structure employed by the Quakers.

No attempts have been made as yet to regularize languages other than English, but at least for the Romance languages it seems on first view that such regularization can be accomplished.

The particular rules of importance to Mechanical Translation are: one word order; one meaning for each word; and one form for each word.

The experience gained in using model language at the Washington Public Opinion Laboratory indicates clearly that regularization of a language minimizes the points brought out by Dr. Bar-Hillel. The discussion showed that the conference was in substantial agreement that regularization by use of the concepts of a model language is feasible and directly applicable to the problems of mechanical translation. In particular, so far as the machines to be employed are concerned, the machine men present felt that it could be a decided advantage in reducing the complexity of equipment required.

Session III - June 18, 1952
Chairman - A. C. Reynolds, Jr.

Prof. Victor A. Oswald, Department of Germanic Languages, University of California, Los Angeles, presented the first paper entitled "Word-by-Word Translation." Prof. Oswald and Dr. Harry D. Huskey, Assistant Director, National Bureau of Standards Institute for Numerical Analysis, University of California, Los Angeles, jointly conducted experiments in the translation of a text in mathematics and another in brain surgery from German into English. The investigation by Dr. Oswald indicated that word-by-word translation from German into English was a virtually impossible task, chiefly because of the fact that German "articles" are also "words." Also, German sentence structure is such that word-by-word translation from German into English becomes virtually meaningless. Initial investigation resulted in a published report entitled, "Proposals for the Mechanical Resolution of German Syntax Patterns."

Although word-by-word translation seemed impossible, breaking of the German sentence into a block-by-block formation, in which each block has a certain specific syntactical function, was far more profitable. Regularization of the German language and other languages of similar structure thus appears to be dependent upon such block-by-block analysis. The "Proposals" indicate that machines can be instructed to recognize syntactic connection upon this basis.

The second major consideration for block-by-block translation is the problem of recognizing and interpreting the meaning-bearing words within a block. Syntactic connections will almost infallibly identify the word function and hence function recognition can be programmed. Linguistic research, particularly that conducted by Prof. William E. Bull, Department of Spanish, University of California, Los Angeles, (also a participant at the conference) shows clearly that the only meaning-bearing forms that can be isolated are nouns, verbs, adjectives, and possibly adverbs. In general, of these classes, nouns are by far the most useful and used bearers of meaning. No system yet proposed will solve the problem of multiple significance of the meaning-bearing words. However, within a specific subject, a meaning-bearing word in general has only one specific meaning. This fact can be utilized to advantage in mechanical translation in which the criterion of meaning is determined by the subject matter being considered. Dr. Oswald proposed to take advantage of this fact by the use of what he termed micro-glossaries. These micro-glossaries would be constructed on the basis of the words most commonly used in specific subjects of interest; one such glossary being constructed for each subject to be translated. Mechanically, this means that two memories would be employed in a machine; one, a most used general vocabulary for the languages being processed; and two, a specific micro-glossary to assign specific meanings to words that would otherwise have a multiplicity of meaning; that is, if all their fields of usage were to be considered simultaneously. The concept of a micro-glossary and the use of block-by-block syntactic recognition in the machine met with favor from all the participants in the conference. The linguists appeared certain that block-by-block syntactic analysis of sentences could be accomplished and likewise were in agreement as to the reduction of ambiguity in the meaning of a word when only one field of interest was to be considered. The engineers present fully recognized the advantage to be gained from the reduction in size of memories growing out of the micro-glossary concept.

Dr. Yehoshua Bar-Hillel presented the next paper on "Operational Syntax." No proposal had yet been presented to the conference regarding a means of programming a machine for recognizing syntactic connections. Dr. Bar-Hillel, examining this problem as a problem in symbolic logic, has discovered certain relationships that exist within the syntax of sen-tence structure. Further, he has discovered that these can be quite readily symbolized in the form of symbolic fractions. A simple multiplication of the fractions, which results in the cancellation of like quantities in the numerator and denominator, results in a unique symbol indicative of the functions of the word block so analyzed. Use of this analysis permits ready recognition of word blocks functioning as nouns, verbs, adjectives, or adverbs.

The identification results in the ability to rearrange the syntax of the source language into the syntax of the target language. This is a simple arithmetic operation that can be readily programmed on a machine. The investigations to date have been preliminary, but indicate that the field is limited only by the number of languages which it would be profitable to so analyze.

This was a completely new concept to the linguists of the conference who had intuitively felt that such a structure did exist but without the tools of symbolic logic had been unable to isolate the essential features that lead to the exceedingly simply arithmetic operations. The engineers immediately recognized the extreme advantages and the simplicity of the computing loops necessary to give the machine the ability to recognize word block functions and programmed reorganization of sentence structure.

Prof. William N. Locke, Department of Modern Languages, M.I.T., presented the third paper on “Mechanical Translation of Printed and Spoken Material.” This paper was presented orally only, no copies having been made for distribution.

Prof. Locke is interested in the potentiality of using voice input to produce either a voice output or a printed output. He drew on work that has been conducted at the Bell Laboratories, at the Haskins Laboratories, at M.I.T., and elsewhere on the analysis of speech and the recognition of the components that form the spoken word. It appears at the present time that 8 such components uniquely determined a sound. Recognition of these 8 elements leads to the identification of one sound to the exclusion of all other sounds. It was Prof. Locke’s contention that a machine could be built to recognize these 8 components and give a unique output (phoneme). The phoneme so constructed could be used with other phonemes to locate a specific unit within the memory whose meaning in the target language would be the same as the meaning in the source language. This of course pre-supposes the utilization of the philosophy in constructing memories as outlined in the previous pages of the conference.

The discussion of Prof. Locke’s paper was completely speculative since devices capable of so analyzing sounds are not yet in existence and it appears that it will be sometime in the future before such an art can become a science.

Session IV - June 19, 1952
Chairman - Dr. A. Don Booth

Dr. Victor A. Oswald presented the first paper, entitled “Microsemantics.” This paper continued the analysis that Dr. Oswald had presented on the preceding day in his discussion of word-by-word translation. He was now concerned with the fact that, in general, editing of the subject material would be required both before translation, in the source language, and after translation, in the target language. The problem is to simplify as much as possible the work required in such pre-editing and post-editing.

Assuming that syntactic considerations could be solved by such an analysis as that proposed by Dr. Bar-Hillel, the work of translation would be very greatly facilitated by the use of special-ized glossaries concerned with the specific sub-ject matter of the material being translated. (Dr. Oswald terms this type of glossary a micro-glossary, and the analysis that leads to it, micro semantical investigation.)

The data obtained from every sort of linguistic frequency count when arranged according to descending numbers forms a monotonic descending curve. The words of highest frequency drop quite abruptly; words of medium frequency start flattening out; and words of highly specialized meaning that are used but seldom cause the curve to approach the horizontal axis asymptotically. The upper segment of the curve contains the words which are usually found in the normal or everyday vocabulary of a language, and contains about 80 per cent of the actual volume of the material. Unfortunately, these terms consist mainly of articles which convey but little meaning; the meaning-bearing forms, and in particular the nouns, are represented by the tail of the curve. All languages exhibit this characteristic curve. Thus, in order to find those words conveying the major meaning in any text, we are concerned with the tail of the curve rather than the large grouping of words occurring at the beginning of the curve. Considering that this particular section of the curve is representative of a micro-glossary of a specific subject in the language, the words of this section in general will have one and only one mean-ing.

To verify this assumption, Dr. Oswald analyzed nearly a hundred papers in German on the subject of brain surgery. Technical nouns were abstracted from the first article. Additional nouns were added from the second article, and so through the complete series of texts em-ployed. Each succeeding text was chosen from a different field of brain surgery. The amazing fact developed that after the fourth article, the glossary derived covered an average of 80 per cent of all the technical nouns in each succeeding article. From this, he constructed a micro-glossary that he considers representative of the field of brain surgery in the German language.

A similar glossary of non-technical nouns was also compiled from the same series of articles. The frequency curve of the non-technical nouns was the same as that of the technical nouns. In other words, the brain surgeons are not only compelled to choose their technical nouns from a limited vocabulary, but their pattern of communication is so limited by practice and convention that even the range of non-technical nouns is predictable.

We may generalize, although perhaps dangerously, that the same phenomenon will appear in all technical fields of a restricted nature.

The micro-glossary was employed in programming translations on the SWAC in cooperation with Dr. Harry D. Huskey, Assistant Director, National Bureau of Standards Institute for Numerical Analysis, University of California, Los Angeles. The translations so obtained conveyed the meaning of the original article with correlations of meaning better than 90 per cent, on the assumption that the problems of syntax and contextual modification had previously been solved. Even without this assumption, the translated articles, when presented to a specialist in the field, in the raw un-edited form, conveyed the major portion of the meaning of the original article in the original language.

The discussion that followed the paper clearly showed that the linguists working in other languages than German were in complete agreement as to the ease with which such micro-glossaries could be constructed. The engineers and scientists, from their knowledge of technical articles in their respective fields, indicated that the size of micro-glossaries in these fields would be as small in comparison to the complete vocabulary of a language as Dr. Oswald postulated. All agreed that the use of such micro-glossaries would enormously reduce the amount of memory required in a translating machine.

In particular, the discussion centered on isolation of nouns as the major meaning-bearing words of a language. A rough analysis was made of the language being used around the table, and it was quite evident that in general verbs employed in conveying meaning through speech are in the present tense and in the vast majority of cases the verb is a form of the verb "to be." Since information is adequately conveyed by speech, it seemed reasonable to the participants that a translation which would ignore tenses and concentrate on nouns which - in newspaper parlance - convey the who, what, when, where, and how, of a statement, would adequately convey to a post-editor the necessary raw material to be employed in producing a polished translation. Dr. Oswald was congratulated by the group for his work and analysis of this phenomena.

Prof. William E. Bull, Department of Spanish, University of California, Los Angeles, presented the second paper entitled "Frequency Problems in Mechanical Translation." Prof. Bull's investigation in Spanish literature paralleled the investigations of Dr. Oswald. Running texts in Spanish literature, which employed a general vocabulary rather than a restricted vocabulary, verify in detail the existence of the same phenomenon in general language as occurred in the restricted field of brain surgery, but Prof. Bull stressed that low frequency, unpredictable terms often carry critically important meaning.

Prof. Bull exhibited numerous slides showing the frequency counts of words, the frequency occurrence of particular parts of speech, and the frequency counts of words within the classi-fication of a particular part of speech. He dis-cussed in some detail the problem of deter-mining syntactic connections in Spanish sen-tences. He also discussed the type of work and the type of personnel required to extend know-ledge in this field not only for Spanish but also for other languages of interest.

Prof. Bull's paper was in part abstracted from a monograph not yet published. Therefore, he did not present a written paper to the participants of the conference, and this material is at present unavailable.

Substantially, Prof. Bull's paper was a verification of the work of Dr. Oswald and indicated the fruitfulness of this approach to the problem of Mechanical Translation. A discussion of the means required to further extend the investigations showed clearly that the analysis could be facilitated by the use of punched cards. Such mechanization can enormously increase our knowledge of language structure, whereas the present handwritten and hand-sorting techniques are far too slow to materially aid in the solution of the problems of mechanical translation. Prof. Bull accepted the suggestion that he investigate the possibilities of employing punched cards as a means of extending the scope of his research.

The third paper was presented by Prof. Erwin Reifler and was entitled "General Mechanical Translation and Universal Grammar." Prof. Reifler has inaugurated a new school of linguistic investigation which is currently known as "Comparative Semantics." Prof. Reifler has been investigating languages in order to discover such patterns of verbally conveying meaning, underlying the actual words and syntax of a language, as are common to all languages. Such a structure could form a "universal grammar."

Mechanical translation poses the following question: "Is it possible to solve the problems of Mechanical Translation in such a way that one and the same preparation of the code text may serve for a Mechanical Translation into many different languages?" The existence of a universal grammar would most assuredly assist in the solution of this problem if such a grammar could be shown to exist. To date, the science of linguistics states that no such universal grammar exists, but linguists do speak of language universals. In particular, many highly interesting cases of parallel development in the evolution of the expression of meaning amongst structurally unrelated languages do exist. The universals may be used to re-adjust the language structure to form what Prof. Reifler terms "adjusted model target languages." This is in line with the recommendation that Prof. Stuart C. Dodd presented in his paper on "Model English." Use of the adjustment clearly simplifies the mechanical translation problem and the engineering required for its solution.

The discussion of the paper reinforced the conclusions of the discussion on Prof. Dodd's paper. It is encouraging to note that where Prof. Dodd has restricted his considerations to English and hypothesized extension to other languages, Prof. Reifler, working from a completely different viewpoint and another purpose in mind, arrived at the same conclusions as to the feasibility of regularizing a language and further demonstrated our ability to regularize major language groups of the world.

Session V - June 20, 1952
Chairman - Prof. Wm. E. Bull

Dr. Harry D. Huskey, Assistant Director, National Bureau of Standards Institute for Numerical Analysis, University of California, Los Angeles, presented the first talk, "Basic Machine Operations in Mechanical Translation." No paper was prepared for distribution to the members of the conference.

Dr. Huskey reviewed the problems encountered in programming German translations in collaboration with Dr. Oswald. The problems encountered were, to a certain extent, peculiar to the SWAC, which was the machine available for the translation. The basic problems were the construction of a vocabulary for entry into the machine, the derivation of a system of addressing to find particular units in the memory, and the syntactic programming to obtain correct sentence structure in the output of the machine. These problems are basic to any machine translation. In general, the design of the machine will govern the type of programming required. The use of two types of memories seems desirable – the first having short access time and the second, which will contain words of infrequent use, having a longer access time. The arithmetic operations required for the construction of the correct sentence structure will be dependent upon the arithmetic devices provided with the machine. The complexity of the machine, if a machine is constructed for the sole purpose of mechanical translation, will be a function of the degree of accuracy required in the translation. This in turn will be dependent upon the allocation of time for pre-editing the material for machine input and post-editing of the machine output.

The second paper was presented by Mr. J. W. Forrester, Director of Digital Computer Laboratory, M.I.T., on the subject of "Problems of Storage and Cost."

This also was presented in the form of a talk, no written material being distributed.

Mr. Forrester presented no cost items that are not known to computers and business machine engineers. His major purpose was to indicate to the linguists present the cost of the machine that they were proposing. Techniques employing magnetic drums, magnetic tapes, and electrostatic storage devices singly and in combination with one another were presented for consideration. The most economical array consists of an intermediate memory and computing unit of low access time and a large scale memory of long access time. The cost of the machine is dependent on the same considerations as listed by Dr. Huskey.

The third paper was presented by Dr. A. Donald Booth, Director, Computation Laboratory, Birkbeck College, London. The title was changed from that listed in the program to "Some Methods of Mechanized Translation," which was written in collaboration with Dr. R. H. Richens of the Biological Laboratories of the University of London. General principles of mechanical translation, as scheduled and programmed on the computer built by Dr. Booth for the University of London, were discussed.

The use of punched card machinery was compared with the use of an automatic digital computer. Time comparisons were worked out that favored the use of the automatic digital computing machinery by a time ratio of at least 7 to 1. Examples of translations in the field of genetics from Albanian, Danish, Dutch, Finnish, French, German, Hungarian, Indonesian, Italian, Latin, Latvian, Norwegian, Polish, Portuguese, Rumanian, Spanish, Swedish, Turkish, Arabic, and Japanese were given. Usable translations in each of these cases, despite the limited storage available with Dr. Booth's computer, were obtained. Post-editing was necessary in all cases, however, to produce a readable, although not necessarily more intelligible translation.

The fourth paper was presented by Prof. Wm. E. Bull and was concerned with the possible future effect of the concept of mechanical translation on the teaching of foreign languages. Prof. Bull stated that the concept of mechanical translation necessitates a completely new approach to the problem of language teaching. An analogy was drawn between a machine into whose memory a vocabulary had not been incorporated and a student into whose brain such a vocabulary must also be introduced. The approach in teaching syntactic connections to both the machine and to the student in terms of the programming required to obtain syntactically correct constructions from the memory storage was discussed. Prof. Bull reached the conclusions that the same considerations that govern the choice of vocabulary and the use of intermediate and large scale memories in the machine could be advantageously incorporated into the teaching of languages as well as the design of machines for mechanical translation.

Dr. Louis N. Ridenour was unfortunately unable to attend the conference, and his paper on "Learning Machines" was not presented.

In his place, Prof. James W. Perry, Research Associate, Center for International Studies, M.I.T., presented a paper on "Machine Techniques for Index Searching and for Machine Translation." This paper was an elaboration of the talk that Prof. Perry presented at the opening public session of the conference. To a considerable extent, the concepts in the paper were based on Prof. Perry's experience in setting up coding and indexing systems for hand-sorted punched cards, and also on his experience with the library-cataloging machine developed by IBM. Fundamentally, the same conclusions as to memory and access times were reached by Prof. Perry as had been previously derived by the other participants in the conference.

Session VI - June 20, 1952
Chairman - Prof. Wm. E. Bull

The closing session of the conference was devoted to a consideration of organization for future research. A seven-man committee was organized at this session to act as coordinators and consultants for further work. The committee is composed of Dr. Yehoshua Bar-Hillel, as chairman; Prof. Leon Dostert, secretary; and Dr. Olaf Helmer, Dr. Harry D. Huskey, Prof. Erwin Reifler, and Mr. A. C. Reynolds, Jr., as members. Dr. A. Donald Booth was placed on the committee as the European representative.

In the organization for future research, the conferees were asked to what degree they were interested in future work and in which areas they wished to participate.

Dr. Booth will continue with the work he has already started with Dr. R. H. Richens at the University of London.

Prof. Bull is interested in the field of linguistic problems of translation and as part of his research activity will continue with his study of the Spanish language. He is not concerned with mechanical translation as such, but recognizes the necessity for, and the value of, his linguistic work in reaching this goal.

Dr. Dodd will continue his work in the studies of regularizing languages and determine the degree of extension possible in languages other than English.

Prof. Dostert intends to work actively, through the Institute of Languages and Linguistics, Georgetown University, in the derivation of principles for the use of machines in translation.

Dr. Olaf Helmer stated that the Rand Corporation is interested from the theoretical viewpoint, but in all probability at the present time will confine itself only to theoretical work as secondary to its work on computers.

Dr. Huskey had no comment other than that he would continue to collaborate with Prof. Oswald.

Prof. Oswald is interested in extending the concept of micro-glossaries and in the study of syntactic relations. He intends to continue work in the programming of translation for machines.

Prof. Reifler is extremely interested in demonstrating the existence of universals in grammar, and in applying these universals to the problem of mechanical translation.

Dr. Bar-Hillel will continue his basic research in symbolic logic and its applications to the field of mechanical translation.

Dr. Jerome B. Wiesner, speaking for the M.I.T. staff present, stated that the research laboratory at M.I.T. is very much interested in the application of computer techniques to the problem of mechanical translation and that if a concrete program was formulated, financial support could quite conceivably be forthcoming from the Research Laboratory.

Mr. Duncan Harkin of the Department of Defense stated that the Department of Defense was vitally interested in this problem and, like Dr. Wiesner, if a concrete proposal for such a translation and subsequent demonstration could be formulated, the Department of Defense would be prepared to give financial backing.

Mr. Reynolds stated that IBM was interested in the application of its present punched card techniques and its computers to this problem and as such would participate on the basis of exchange of theoretical information with the members of the conference.

The conference closed on a note of optimism regarding the potentialities now known to be physically present in the concept of mechanical translation.

---

NEWS (cont. from p. 37)

issue for 1954 (Vol. 20, pp. 259-351) to problems of translation. Although written primarily from the point of view of the descriptive linguist, the issue contains several articles in which problems of mechanical translation are taken into account. The most important among these are Z. Harris' "Transfer Grammar" and C. F. Vogelin's "Multiple Stage Translation."

MECHANICAL TRANSLATION DISCUSSED AT MEETING OF THE AMERICAN CHEMICAL SOCIETY

At the American Chemical Society meeting in New York the Division of Chemical Literature held a symposium on aids to the use of the foreign chemical literature, September 14. Of the thirteen papers presented, three were on mechanical translation. W. N. Locke spoke on "Autotranslation: Development and Prospects." James W. Perry and Anthony Oettinger presented "Practical Goals for Machine Translation" and L. E. Dostert contributed "Characteristics of Recent Mechanical Translation Experiments." The room was well filled with an estimated 120 persons. In the evening a demonstration was given by the International Business Machines Company of their model 701 general purpose computer programmed to translate sample Russian sentences. The routine was devised jointly by IBM and Georgetown University.

A. D. BOOTH of Birkbeck College, London, writes that Mr. J. Cleave has been appointed by the Department of Scientific and Industrial Research to do research work with him in MT.

THE editors would like to express appreciation to those who have sent in articles, comments, news items, and bibliography items. It is only through your cooperation that MT can achieve its full value as a medium of communication in the field.

Victor H. Yngve 46
The Machine and the Man
Mechanical Translation, Vol. 2, No. 2, pp. 20-22 (August 1954)

A discussion of the creation of new positions which the perfecting of a mechanical translating machine will afford. These positions entail more than design, construction, maintenance, clerical work and the need for man to provide a program or "operational syntax" for the machine. There will be a need for someone to note the machine's shortcomings and to adjust the original tentative program accordingly, thereby "training" the machine to learn new material, profiting from its own mistakes. Availability of machine-produced translations, adequate for

Yngve – The Machine and the Man (Cont.)

many purposes, will increase the demand for more accurate human-produced translations.

Jane Grace

Erwin Reifler 47
Report on the First Conference on Mechanical Translation
Mechanical Translation, Vol. 2, No. 2, pp. 23-32 (August 1954)

A report on the first meeting of linguists and electronic engineers interested in the problem of mechanical translation, held at M.I.T. in June 1952. Professor Erwin Reifler, one of the linguists participating in the three-day conference, discusses the various papers that were presented. Abstracts of most of these papers may be found in MT, Volume 1, Number 1.

Jane Grace

Robert Leon Gourdon 48
"Verfahren und Vorrichtung zur Selbsttätigen Vollständigen und Augenblicklichen Übersetzung von Schriftstücken in Verschiedene Sprachen"
German Patent 911187, May 10, 1954

German patent covering a mechanical and optical device which is claimed to be capable of providing a word for word translation. The device uses a keyboard input and photographic film for storage and output.

V.H.Y.