MECHANICAL TRANSLATION

DEVOTED TO THE TRANSLATION Of LANGUAGES WITH THE AID OF MACHINES

VOLUME TWO, NUMBER THREE DECEMBER, NINETEEN FIFTY FIVE

News

ACLS COMMITTEE

A committee has been formed by the American Council of Learned Societies for the purpose of following progress in mechanical translation and the application of mechanical methods to linguistics in general. The members of the committee are Norman A. McQuown, chairman, Paul Garvin, Martin Joos, and W.F. Twaddell.

MEETING IN CLEVELAND

A number of people who have been active in the field of mechanical translation met in Cleveland January 18. The occasion was a Conference on the Practical Utilization of Recorded Knowledge held under the auspices of the School of Library Science, Western Reserve University. One of the evening meetings was devoted exclusively to mechanical translation. Reviews of the various programs under way were given and there was extensive discussion of future possibilities and plans. The meeting was chaired by Major M.H. Wienberg and was attended by Y. Bar-Hillel, Lt. L.F. Buckland, L.E. Dostert, K.E. Harper, W.R. Hill, A. Koutsoudas, L.R. Micklesen, J.W. Perry, E. Reifler, G. Shiner, V.H. Yngve, as well as representatives of various interested segments of government and industry.

MT IN MOSCOW

We were glad to hear of mechanical translation activity in the Soviet Union. According to newspaper accounts, there is interest in MT by people associated with a large computer at the Institute of Exact Mechanics and Computing Technique of the Soviet Academy of Sciences, Moscow.

The reports indicate that mechanical translation routines have been devised for this computer. Sergei A. Lebedev, the institute's director, is quoted as saying, "The word-for-word translation thus obtained is edited by the machine with an eye to the grammar and syntax of the language."

See Pravda, December 4, 1955 and January 22, 1956, also The New York Times, December 11, 1955 and The Christian Science Monitor, March 21, 1956. See also abstract 64 in this issue.

APPOINTMENT

M.I.T. has announced the appointment of William N. Locke to the post of Director of Libraries. Professor Locke will continue to head the Modern Language Department and will continue his interests in mechanical translation.

Braille Transcription and Mechanical Translation

John P. Cleave, Birkbeck College, University of London, London, England

TRANSCRIBING romanized print into Braille suitable for reading by the blind is a problem which has similarities to those arising in mechanical translation. The theoretical problem of mechanical translation is to construct an operational syntax - a set of formal rules of translation prescribing operations to be performed on the text to get the output text - entirely in terms of patterns of input words and types of words and such information as may be contained in the dictionary. And this is simplified already, firstly by the small vocabulary (consisting of a definite number of letters, capitalized letters, punctuation marks, etc.) and the absence of ambiguity and, above all, the existence of explicit rules for transcription which are already partly formalized.

The Braille Systems

Braille is a system of embossed characters formed by six dots arranged and numbered as in Fig.l(a). In the project outlined here the output of the computer presents the Braille characters as a series of six "1's" or "0's" corresponding to the six Braille dots. Thus the Braille character of Fig.l(b) is represented by the binary number of l(c).

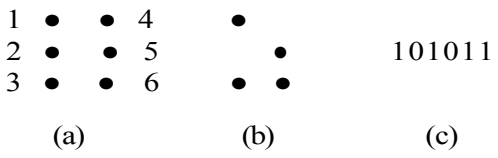

Figure 1

While to each letter-press character there corresponds one Braille sign, there are Braille characters (single-cell contractions) and pairs of Braille characters (double-cell contractions) which under various conditions represent groups of inkprint letters. Thus, the Braille character of Fig.2 represents the group "wh" in that order.

The rules of Braille largely concern the conditions under which contractions can be made. There are four grades of Braille: Grade I, uncontracted; Grade "one-and-a-half"; Grade II, moderately contracted; Grade III, highly contracted. The latter grade is rarely used. Grade I presents no problem to the computer. Grades "one-and-a-half" and II are the more profitable lines of inquiry,

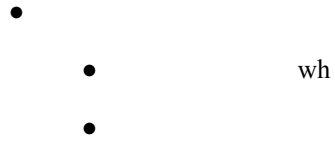

Figure 2

The problem to be dealt with is that of constructing a program by which an electronic computer will do the work of making the contractions correctly. We envisage an input organ to the electronic computer with a keyboard with keys for all the characters used in inkprint (including punctuation marks). The output from this organ is in the form of binary numbers (machine characters) on which the computer operates and finally obtains from each such number a six digit binary number representing the six Braille dots. (Fig.l) An output mechanism, similar to an ordinary teleprinter (it could in fact be such a piece of equipment fitted with a mechanical device), will convert this number into the Braille characters as actually used.

The Braille signs used in this project are as shown in Fig.3. These characters are divided into classes called "lines." Line 1 is formed by dots 1-2-4-5. Line 2 is formed by adding dot 3 to each of the characters of line 1, and line 3 by the addition of dots 3 and 6 to line 4. Line 4 is formed by the addition of dot 6 to line 1 signs. Line 5 is obtained by repeating line 1 in a lower position. This classification has no significance as far as the Braille rules are concerned.

A further classification of Braille signs, which cuts across the "line" division, is the classification into "lower signs" and "non-lower signs"; a lower sign is a Braille sign which does not

contain dot 1 or dot 4. The lower signs are all those of line 5 together with "com" of line 6. This again is a formal property of the Braille

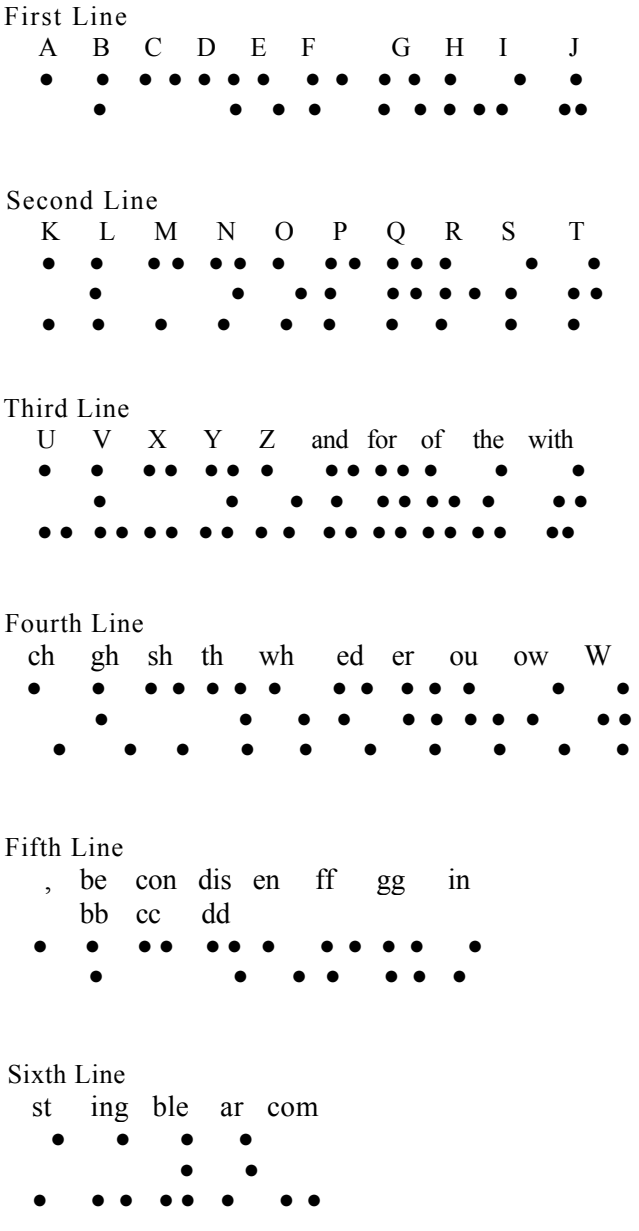

Figure 3

sign, but for technical convenience it is explicitly represented by a code digit attached to the coded Braille. The rule concerning the contraction of double letters requires explicit mention of the lower sign property.

Formalization of the Rules

The rules followed in this work are those printed in Standard English Braille. [1] The rules as expressed in the bookle.t are not all usable for a mechanical transcription of inkprint characters into Braille as they stand, though they are perfectly satisfactory for a human agent. To be put in a form suitable for the construction of a machine program the rules must be formalized. That is, all reference to terms which cannot be given an extensional definition in terms of the machine characters, or a definition in terms of their formal properties, must be eliminated. For instance, rule 34 reads:

> Contractions forming parts of words should not be used when they are likely to lead to obscurity in recognition or pronunciation and therefore they should not overlap well-defined syllable divisions. Word signs should be used sparingly in the middle of words unless they form distinct syllables. Special care should be taken to avoid undue con - traction of words of relatively infrequent occurrence.

The principal term in this rule is "syllable." It would be possible to formalize this term if a complete list of syllables could be compiled. This would be a clumsy procedure and would require comparison of incoming words with a large dictionary for recognition of syllables. Similar difficulties arise with "pronunciation," though the problem is largely solved when the "syllable" question has been resolved. The most simple way to resolve the issue is to ignore the restrictions imposed by this rule. Another, which includes a non-formal restriction, is rule 21:

> The word signs and, for, of, the, with, a, may follow one another without a space where the sense permits. . .

The condition "where the sense permits" is impossible to formalize fully except by constructing a list of phrases in which the elimination of the space between these "and-words" may be effected without destroying the sense. However, the sense may not be determined by the phrase but by the whole sentence. The task of including this condition in its entirety in a machine program is now immense. Confusion could arise when a space is eliminated between and-words where at least one is part of a word.

1 Published by the "National Institute for the Blind," London, 1932.

The restriction could then be formalized to read:

> ". . unless at least one of the and-words is part of a word"

It is simpler to ignore the wide restriction and to base the space-elimination entirely upon the occurrence of the words. More will be said of this rule later.

On the other hand some of the rules are already adequately formalized. For instance, Rule 27:

> The contractions bb, cc, dd, ff, gg, may only be used when they occur between letters and signs of the same line of Braille.

Since "word" and "line" can be given formal definitions the rule as it stands is sufficient though it is more explicit (ignoring the complication caused by "line") if we simply say:

> Use the contractions bb, cc. dd, ff, gg if the sign preceding and the sign following b b, c c, d d, f f, g g are neither spaces nor punctuation marks.

An important principle in formalizing the rules is the explicit representation in the machine characters of the properties used for the operation of the program. For instance, a word can be defined formally as the series of signs lying between signs each of which is either a space or punctuation mark. We therefore require that the computer recognize the punctuation marks. It would obviously be possible to define the punctuation marks extensionally as "either the comma or full stop or exclamation mark or.." The process by which the machine recognizes the punctuation mark is then quite complicated, involving comparison of the incoming letter with each punctuation mark in turn, which is slow and wasteful of storage space. The simplest procedure is to indicate membership of this group of words by a digit of the machine character. Several other properties, either of the Braille characters or the letter-press characters, and membership of various other classes are best represented by digits of the machine characters.

The Structure of the Machine Characters

The machine characters must bear the six digits representing the Braille dots. It is technically convenient to represent the membership of the various classes of sign by a set of three digits (the code-digits) preceding the six Braille digits, so that the machine character is a number with nine binary digits. Thus the machine character has the following structure:

1st position punctuation digit
2nd position "and"word digit
3rd position "lower sign" digit

These are the code digits. The 4th – 9th positions represent the Braille dots: these digits are the machine representation of Braille.

The first digit, showing whether the letter is a punctuation mark, presents explicitly a property of the alphabetic letter rather than of the structure of the corresponding Braille sign, for a Braille sign may be used either as a contraction or as a punctuation mark (see the signs of line 5). Since some of the Braille rules concern the occurrence of punctuation marks, it is necessary that the machine characters corresponding to such signs carry that information explicitly. Thus the machine can determine the presence of a punctuation mark in the accumulator by shifting left one place and then using the conditional transfer order to discriminate on the sign digit.

Pattern Sensing

A method of detecting patterns of signs is to delay the final printing while sending the last several characters in turn through a series of memory locations. The context of any machine character can then be searched. An illustration of this process is provided by the following method of operating Rule 21 mentioned above. The series of machine characters, after having been modified by the contraction program to produce the and-word characters, is sent serially through five memory locations. If the conditions for space elimination are not present, the character in the fifth position is sent to the "print routine" which removes the code digits and prints the six digits representing the Braille sign. The characters in the remaining positions are then shifted one place by the "shift routine" leaving the first place to be occupied by a new character from the contraction routine. Rule 21 in the form required by the machine program now reads:

> (i)if there are either punctuation marks or spaces in locations (1) and (5) go to (ii); if not go to the print routine.
> (ii) if there is a space in (3) go to instruction (3); if not go to the print routine.
> (iii) if there are and-words in both positions (2) and (4) shift the character in (2) to (3) and that in (1) to (2) (space-elimination); if not go to the print routine.

This version of the rule is in fact weaker than the original since it permits only pair-wise juxtaposition of "and"words. But it does deal ade-

quately with the majority of cases. It would be possible to construct a routine for effecting the space-elimination in all the circumstances demanded by the formalized version:

"the 'and' words may follow one another without a space unless at least one of them is part of a word"

This, however, would be rather long and would not be justified by the frequency with which three or more consecutive and-words occur, compared with the relatively large frequency of pairs of and-words.

More complicated procedures of a similar nature are necessary to operate the rules concerning numerical expressions, ellipsis, compound lower signs and capital letters.

The Dictionary

In Grade 'one-and-a-half' it is unnecessary to have a dictionary for the contractions; incoming letters may be compared on arrival with possible members of contractions by means of a "contraction routine." Thus, if an "a" is detected, the contraction routine compares the following character with "r". If an "r" is found, the "ar" contraction is subjected to the next part of the program; if not, "a" is sent to the next part of the program after which the letter following "a" is examined to determine whether it could be the initial letter of a group which could be contracted.

Grade II Braille, on the contrary, contains so many contractions that it is necessary to use a "dictionary" of groups which can be contracted. Characters must then be fed in serially and stored in a set of temporary locations - the Initial Word Store - until a whole word has been received. The dictionary matching mechanism then takes the first letter in the Initial Word Store and finds the longest dictionary entry which is part of that word. The appropriate contraction is selected and sent to another set of storage locations - the Final Word Store - after which the remainder of the word is treated in the same way. Should no entry be found, the first letter is sent to the Final Word Store and the matching procedure started with the second letter.

There may be several ways of contracting a word. The choice between the methods of contraction is governed by considerations of length. That way must be chosen which gives the shortest transcription. The case where two different methods of contraction yield words of equal length is governed by rule 35:

> In cases where a word may according to the above rules be contracted in two or more ways, each saving the same amount of space, that way should be selected which produces the most readable combination of dots. If the same space is saved, simple contractions are better than two-celled word-signs. Avoid using Double Letter Signs where there is an alternative single cell contraction.

The dictionary is so constructed that the shortest set of contractions is automatically chosen. For instance, "themselves" precedes "the" in the dictionary so that if "themselves" occurs in the Initial Word Store it is compared with the appropriate entry before being compared with "the". If, however, "them" occurs in the text, the longest dictionary entry occurring which is part of that word is "the". The priority rule for single-cell contractions is solved by including in the dictionary those phrases which provide a double-"translation." For instance, the phrase "oner" occurs in the dictionary and precedes "one". "Oner" may be contracted in two ways - "one r" and "o n er. "In the first case "one" is a two-cell contraction so that "one r"occupies three cells. In the second case the translation occupies three cells since "er" is a single-cell contraction. By rule 35 "o n er" is the correct translation of "oner" so the dictionary includes o n er as the dictionary entry. Thus, Rule 35 does not appear explicitly in the machine program but is implicit in the construction of the whole program and, in particular, of the dictionary.

The Signal System in Interlingua — A Factor in Mechanical Translation

Alexander Gode, Science Service, Division de Interlingua, New York City

PREOCCUPATION with the problems of communication across language barriers has resulted in our time in the perfection, progressive application, or investigation of three new techniques. They are (1) simultaneous translation, as practiced in the United Nations General Assembly, and more and more extensively also in the most varied international congresses of scientists and other groups of specialized endeavor, (2) Interlingua, as utilized currently (especially for medical summaries) in increasing numbers of scientific periodicals and printed programs of international congresses, and (3) mechanical translation by electronic computers, as envisaged especially by scholars at Georgetown University, the Massachusetts Institute of Technology, the Universities of California and Washington, and Birkbeck College (University of London).

The technique of simultaneous translation seems definitely established. It is to be assumed that it can be further perfected by technological improvements of the machines with which it operates and also by additional refinements in the training of the translating personnel. The psychological processes which the simultaneous translator experiences--his tendency to identify himself with the orator after the fashion of a good actor, his "schizophrenic" endeavor to be simultaneously, so to speak, a listening Russian balalaika and a resonant Spanish guitar, and so forth--make fascinating material for descriptive and analytical studies in some branch of metalinguistics. But no fundamental research seems required at this time to lift the technique of simultaneous translation into the saddle or to keep it there.

There is likewise very little to be done to extend the limits of the potential usefulness of simultaneous translation. These limits are completely clear, even if they represent a suit of armor the little giant must grow a lot to fill out completely. Simultaneous translation is a technique that can be applied wherever the spoken word in one particular language needs to be understood for immediate reaction by groups of individuals whose language masteries do not cover that one particular language. The product of simultaneous translation need never and can never serve as an "official" rendering ready and valid for incorporation in a permanent printed or otherwise published record.

The relation of Interlingua to the technique of simultaneous translation is on the whole one of irrelation. There can be no competition between the two but only peaceful coexistence and -- let us hope -- cooperation. An ideal example of mutual complementation of these two techniques is that of the forthcoming Sixth Congress of the International Society of Hematology. This Congress has announced its selection of English, Spanish, French, and Interlingua as official languages. All publications -- announcements, programs, etc. -- which the Congress sees fit to publish will be in either English and Interlingua or in Interlingua alone. All papers presented at the Congress will be read in English, French, or Spanish with simultaneous translation being provided for these three languages.

We may, if we wish, dream of a further sim - plification of this already highly efficient setup. Instead of three languages admitted for the presentation of papers, we may wish for an ultimate liberalization shedding all restrictions. Under such a fantastically ideal setup, participants in the Fiftieth or Seventy-Fifth International Hematological Congress may present their papers in Dutch, Hindustani, Japanese, Hungarian, Finnish, Marathi . . . or any other language of their choosing, with no chaos resulting, thanks to the technique of simultaneous translation and its provision of Interlingua versions of every individual contribution.

Those who cannot appreciate the well-nigh ideal efficiency of these two plans (with the second thrown in to placate incorrigible pursuers of the as yet unlikely), those, i.e., who hold that no international congress can be said to have handled the language problem efficiently if it does not provide for one common auxiliary language for all participants, are looking at things from a viewpoint that is alien to me. I do not wish to criticize them, but I also do not wish to be associated or confused with them.

If it seems impossible to construe any sort of

potential, let alone real, rivalry between the techniques of simultaneous translation and Interlingua, the same obtains to an even more striking degree for the techniques of simultaneous and electronic translation.

The suggestion that in some distant future it may become possible to replace the simultaneous translator by an electronic computer which scrutinizes, analyzes, and then translates the spoken word, belongs as much in the realm of science fiction as does the idea of one universal language used by all mankind in all transnational forms of communication.

The juxtaposition of the techniques of simultaneous translation and mechanical translation by electronic computers reveals one striking difference. The simultaneous translator cannot and must not attempt to analyze. If he parses the statement he hears and then proceeds to search for the best word, he is lost. He must grasp the orator's statements as phraseological Gestalten and react by the spontaneous production of corresponding Gestalten in the target language assigned to him. If there are intervening processes of analysis and resynthesis in the translator's mind, the translator himself must not be aware of them any more than a healthy diner is aware of what happens to a slice of steak on its way to generating a new supply of red blood corpuscles. In contrast to all this the technique of mechanical translation presupposes the most careful and the most detailed analysis.

We know -- and admit or emphasize -- that translation is an art. But much of it can be reduced to simple equations which can be recorded for future reference and which are always correct. How much of it?

In this question, I believe, we have in a nutshell the whole problem of mechanical translation. If not only much of the process of translation but all of it could be reduced to verifiable and ever-valid equations, the linguistic side of the problem of mechanical translation would be solved; nothing would remain open and pending but the construction of a mass-producible, economical machine. I am sure, if Mr. Thomas Watson should fail to avail himself of this opportunity to make his name completely immortal, the late Mr. Henry Ford would step into the breach. All of which is meant to point up the fact that in the whole realm of mechanical translation the engineer is ready while the linguist is not.

It may smack of prejudice if I insert at this point the opinion that the complete reduction of the process of translation to objectively valid equations seems impossible. If we consider the simple German statement, "Ich gehe in die Stadt," all the parts that are needed for an English translation can be covered by simple equations, and even the rendering of this particular "in" as 'into' obeys the objective demand of the following accusative, "die Stadt." I could very well expect a machine to render the passage correctly as, 'I go into the city." By a slightly more complex system of equations I could also expect a machine to translate, equally correctly, 'I go downtown. ' But I do not see how a machine can recognize which of the two, 'I go into the city' or 'I go downtown,' is to be preferred in a given context. Or take an example culled from an actual medical text, where the American author speaks of 'atrial fibrillation and flutter.' Let us suppose we want a machine to translate this passage into French. There are but four equations involved which yield the French words, "auriculaire, fibrillation, et, flutter." And yet I do not see how a machine can decide whether the correct translation is "fibrillation auriculaire et flutter" or rather "fibrillation et flutter auriculaires."

The point of interest here is not that we have proved that machine translation must break down under certain conditions. Perhaps we haven't. The point of interest is that we think we have demonstrated such a breakdown because we cannot find an objective and unambiguous indicator or signal which decides that 'I go downtown' is correct and not 'I go into the city,' that "fibrillation et flutter auriculaires" is right and "fibrillation auriculaire et flutter" wrong.

In lieu of our earlier reference to the process of translation as a complex activity reducible with or without a remnant to objective equations, we might say that the process of translation amounts to making, in the target language, statements which heed all the signals appearing in whatever we or a machine are trying to translate.

In this global conception any spoken or written passage consists of signals, nothing but signals. These might be classified as semantic signals ('cheese' is a semantic signal which suggests the entity 'putrid milk'), intonation signals (depending on its intonation, 'no' may signal surprise, incredulity, or rejection), grammatical signals ("die" in "Ich gehe in die Stadt" signals a rela - tionship of movement into the following entity), etc.

For more conventional purposes it may be better to restrict the meaning of the term 'signal' more or less to what I have just designated as 'grammatical signals.' In the present context, however, we need the more comprehensive interpretation. It permits us to expand an earlier allusion and define the task of the researcher in mechanical translation as amounting to the elabo-

ration of a system whereby all the elements appearing in the finished translation are unambiguously derived from objectively recognizable signals in the original.

This approach permits a type of experimentation which brings out two important principles. First, the system of signals in any given language can be described as consisting of various categories. Second, the refinement of signal categories and sub-categories that need to be considered in a given translation problem depends on the relation between the signal systems of the departure language and the target language. In other words In a given pair of languages that are to be interrelated by the process of mechanical translation, the categories of signals need not be exhaustive. If we interrelate two languages by such a process of translation, we can stipulate experimentally that we want to heed only a specific set of signals, The result of the translation effort can then be criticized in order to determine whether it could be improved by heeding additional signal categories, how far the heeding of ever subtler categories can perfect the finished translation, and whether there is really a remnant of indispensable elements which the target language requires but which cannot be inferred from objective signals.

I present a sample translation from Interlingua into English in which in addition to all semantic equations only the signals for tenses, participles, and plurals are heeded. The passage was chosen at random and happens to be the author's summary of a medical paper.

A Study in Vitro of Serum Antileukemic

1. Was prepared in rabbits a antiserum anti leukemia lymphogenous induced in mouses of the stirps DBA_2, containing antibody against antigens lymphocytic normal and leukemic according to determinations by the test of fixation of complement of Thornton et al.

2. When this antiserum was incubated with antigen lymphocytic normal, all its activity complement-fixing was eliminated except it which reacted with tissue leukemic. It seem that a antibody or a group of antibodys was produced that was specific for this leukemia.

3. A antiserum anti a leukemia lymphogenous induced in mouses of the stirps DBA_2 not itself showed capable, so much in administrations prophylactic as also therapeutic, to protect to degrees significant other mouses of the stirps DBA_2. the which had received inoculations of leukemia transplantable of the same type cellular.

4. Is reported the failure of essays to in - duce leukemia in young mouses feminine of the stirps DBA_2 by paint them with 20-methylcholanthrene in benzene.

The critique of this translation will suggest a few improvements of word choice ('strain' instead of 'stirps,' 'female' instead of 'feminine,' etc.); it will demand correct irregular plurals ('mice' instead of 'mouses') and the use of the lonely personal ending in the third person of the present tense; and finally it will point out as the only major weakness of the translation the un-English position of the adjective which overshadows all other blemishes (including the single instance of a misplaced reflexive pronoun).

If we edit the translation in accordance with these observations (taking only one or two additional liberties of minor significance), we obtain the following version.

An in Vitro Study of Antileukemic Serum

1. We prepared in rabbits an antiserum to lymphogenous leukemia induced in mice of the strain DBA_2, containing antibody against normal and leukemic lymphocytic antigens in ac - cordance with determinations by the complement fixation test of Thornton et al.

2. When this antiserum was incubated with normal lymphocytic antigen, all its complement-fixing activity was eliminated except that which reacted with leukemic tissue. It seems that an antibody or a group of antibodies was produced that was specific for this leukemia.

3. An antiserum to lymphogenous leukemia induced in mice of the strain DBA_2 did not show itself able, either in prophylactic or therapeutic administrations, to protect to a significant degree other mice of the strain DBA_2 which had received inoculations of transplantable leukemia of the same cellular type.

4. We report the failure of attempts to induce leukemia in young female mice of the strain DBA_2 by painting them with 20-methylcholanthrene in benzene.

Aside from the question as to how much of the editing performed on the above piece could be reduced to mechanical reactions to signals in the original, there is also the question whether the comparatively satisfactory result was not possibly due to a very high degree of kinship between the two languages involved, i.e., between Interlingua and English.

There can be no doubt about the closeness of the kinship of Interlingua and English. But this kinship is not exclusive; it is a consequence less of the nature of English than of Interlingua.

I suspect that many of the implications and conclusions of the present survey would collapse or would have to be modified if it were extended to the Slavonic and further to non-European languages. That extension (and modification of my doctrine) I have to leave to others who are qualified to tackle the problem. Left to my own devices, I can merely claim that an experimental juxtaposition of Interlingua with any other European language, carried through after the foregoing Interlingua-English model, would yield the same type of result.

The first passage of our Interlingua text, mechanically translated into German, would read:

War (wurde) bereitet in Kaninchen ein Antiserum anti Leukämie lymphogen induziert in Maus (pl.) von d- (der, die, etc.) Stamm DBA_2, enthaltend Gegensubstanz gegen Antigen (pl.) lymphozytisch normal und leukämisch gemäss Bestimmung (pl.) durch d- (der, die, etc.) Test von Fixierung von Komplement von Thornton et al.

The same passage in French would read:

Etais (était, etc.) préparé en lapins un antisérum anti leucémie lymphogène induit en souris de l- (le, la, les) race DBA_2, contenant anticorps contre antigènes lymphocytique normal et leucémique selon déterminations par 1- (le, la, les) test de fixation de complément de Thornton et al.

This French and the preceding German, no less than the fuller English sample, are definitely editable, i.e., if we suppose that a mechanical-translation setup could produce such texts on a large scale, MT (as the experts call mechanical translation) would be in business. One feels tempted to assert that in the presence of an output of such quality, the question of whether a more refined heeding of existing signals can improve the output, or perhaps actually make it perfect, recedes to a fairly academic background.

In any event, the explanation of the comparatively high quality of our results lies in the specific character of the signal system of Interlingua in relation to that of English, French, German, etc.

It should be possible to dramatize this assertion experimentally by a mechanical translation interconnecting German, French, English or other languages not including Interlingua. Let us use for this purpose the English text on which the above-used Interlingua passage was based. The Interlingua passage itself may here be inserted for the sake of completeness.

Un Studio in Vitro de Sero Antileucemic

1. Esseva preparate in conilios un antisero anti leucemia lymphogene inducite in muses del stirpe DBA_2, continente anticorpore contra antigenos lymphocytic normal e leucemic secundo determinationes per le test de fixation de complemento de Thornton et al.

2. Quando iste antisero esseva incubate con antigeno lymphocytic normal, omne su activitate complemento-fixante esseva eliminate excepte illo que reageva con texito leucemic. Il pare que un anticorpore o un gruppo de anticorpores esseva producite que esseva specific pro iste leucemia.

3. Un antisero anti un leucemia lymphogene inducite in muses del stirpe DBA_2 non se monstrava capace, tanto in administrationes prophylactic como etiam therapeutic, a proteger a grades significative altere muses del stirpe DBA_2 le quales habeva recipite inoculationes de leucemia transplantabile del mesme typo cellular.

4. Es reportate le fallimento de essayos a inducer leucemia in juvene muses feminin del stirpe DBA_2 per pinger los con 20-methylcholanthrena in benzina.

This Interlingua passage was obtained by the devices of human, i.e., non-mechanical translation from an English original which read:

An in Vitro Study of Antileukemic Serum

1. A rabbit anti-DBA_2-mouse-induced lymphogenous leukemia serum was prepared that contained antibodies to normal lymphocytic and to leukemic lymphocytic antigens, as determined by the complement fixation test of Thornton and his associates.

2. When this antiserum was incubated with normal lymphocytic antigen, all of its complement-fixing activity was removed except that which reacted with the leukemic tissue. It appears that an antibody or group of antibodies was produced which was specific for this leukemia.

3. An antiserum to lymphogenous leukemia induced in DBA_2 mice, given prophylactically or therapeutically, did not significantly protect other DBA_2 mice that had been inoculated with a transplantable leukemia of the same cell type.

4. The failure to induce leukemia in young DBA_2 female mice by painting them with 20-methylcholanthrene in benzene is reported.

In putting this passage mechanically into French or German, our interest is to see whether the product is editable as the corresponding product based on Interlingua was editable. The French result is as follows:

Un(e) in vitro étude de antileucémique serum

1. Un(e) lapin anti-DBA_2-souri-induit lymphogène leucémie sérum étais (était, etc.) préparais (préparait, etc., préparé) que contenais (contenait, etc., contenu) anticorps à normal lymphocytique et à leucémique lymphocytique antigène, comme déterminais (déterminait, etc., déterminé) par le complément fixation test de Thornton et son (sa, ses) associés.

2. Quand ce (cet, etc.) antisérum étais (était, etc.) incubé avec normal lymphocytique antigène, tout (tous) son (sa, ses) complément fixant activité étais (était, etc.) éloignais (éloignait, etc., éloigné) excepté que (celui-là, etc.) que (celui-là, etc.) réag-ais (réag-ait, etc., réag-é) avec le (la, les) leucémique tissu. Il apparaît- que un(e) anticorps ou groupe de anticorps étais (était, etc.) produi-ais (produi-ait, etc., produit) que étais (était, etc.) spécifique pour ce (cet,etc.) leucémie.

3. Un(e) antisérum à lymphogénique leucémie induit en DBA_2 souris, donne prophylactiquement ou thérapeutiquement, fai-ais (fai-ait, etc.) ne-pas protéger autre DBA_2 souris que (celui-là, etc.) av-ais (av-ait) été inoculais (inoculait, etc., inoculé) avec transplantable leucémie de le (la, les) même cellule type.

4. Le (la, les) faillite à induire leucémie en jeune DBA_2 féminin souris par colorant les (eux) avec 20-methylcholanthrène en benzine est rapporté.

There are in this string of French words certain sequences that might make sense to an editor of good will. But there are others that cannot possibly be parsed by anyone unless he knows English, and knows it at least well enough to tackle the translation without mechanical help in the first place. The impression left by the corresponding German product is not much better. In lieu of the complete text, this sample may illustrate the point:

Ein (eine, etc.) Kaninchen anti-DBA_2-Maus-induziert(e) lymphogen Leukämie Serum war (wurde) bereitet(e) dass (das, der, etc.) enthalt-et(e) Gegensubstanz zu normal lymphozytisch und zu leukämisch lymphozytisch Antigen (pl.), wie bestimmt(e) durch d-(der, die, etc.) Komplement Fixierung Test von Thornton und sein-Sozius (pl.).

While the samples of English, German, and French evolved by mechanical translation from an Interlingua starting point were so eminently comprehensible and readily editable that a refinement of the mechanical process lost at least some of its urgency, the German and French samples evolved from an English base are at least in part so eminently incomprehensible and uneditable that an immediate identification of the responsible factors becomes imperative. Let us take up at least one representative case.

What could we do to eliminate or reduce the utter confusion of "un lapin anti-DBA_2-souri-induit lymphogène leucémique sérum" and "ein Kaninchen anti-DBA_2-Maus-induziert lymphogen Leukämie Serum"? What additional signals could we have heeded in the English original, 'a rabbit anti-DBA_2-mouse-induced lymphogenous leukemia serum' ? More specifically: What signals are there to decide whether this is a 'leukemia serum' which happens to be 'lymphogenous' or a 'serum' of 'lymphogenous leukemia'? Whether it is a 'rabbit leukemia' or a 'rabbit serum'? Whether it is a 'serum induced against DBA_2 mice' or a 'leukemia induced by anti-mice'? The fact of the matter is that there are no signals to answer these silly questions and quite a few other less silly ones. The English passage is not grammatically comprehensible to anyone not specially prepared by information about the subject matter.

The passage under discussion may be extreme, but it is certainly not unrepresentative. English is rich in unsignaled relationships of a peculiarly complex kind. But, the presence of unsignaled relationships in English or in any other language is not especially noteworthy. It is rather the absence of such relationships that would be news, and incredible news to boot. Signalwise, snow - man, milkman, pitman are quite alike, yet we know that a pitman is not a man made of pits; a snowman is not a man who sells snow; and a milkman is not a man who does his work submerged in milk, even though we have to gather that knowledge from experience not reflected in the corresponding word forms. Signalwise, "Ich gefalle ihm" and "Ich folge ihm" are quite alike, yet we know that the first statement involves a reaction on his part, the second an action on my part although there is no objective signal to mark this difference.

What is important from the point of view of translation and of mechanical translation in particular is not that the signal system of departure language and target language be complete in any absolute sense of the term but rather that they

be compatible. If the departure language supplies signals for categories which the target language does not and cannot represent by special forms and leaves unsignaled other categories which the target language requires, the translation becomes correspondingly more difficult and may even turn out to be impossible.

In the case of the languages used for illustrative purposes in this paper such difficulties are not insuperable but they are quite real. In evolving texts in any of these languages from Interlingua, however, they are all but non-existent. The reason for this seemingly surprising observation is not hard to find. The categories formally signaled in Interlingua are those and only those which the languages summarized in it have in common. If one of the base languages of Interlingua dispenses with a particular signal category, so does Interlingua. If we translate mechanically from Interlingua into English or French or German or any other language of the same general group, we find of necessity that (1) Interlingua gives no signals which our target language finds it impossible to utilize and (2) Interlingua's failure to supply signals of various types customarily present in the target language is restricted to instances where comprehensibility and hence editability is not impaired.

An English text which never signals the difference between ordinary and progressive tenses may sound queer, but it is comprehensible and editable. The same goes for a German text which never signals the difference between a pronoun that refers to "der Tisch" and one that refers to "die Uhr" or "das Buch". And exactly the same, too, goes for a French text which never signals by a verbal ending whether the first, second, or third person is meant.

It is true that many of the specialists of me - chanical-translation research are not satisfied with editable products. They evidently must have arguments which defeat everything I have said to show that there are translation situations in which mechanical processes cannot possibly yield editable results, let alone results that require no editing whatever. Yet these men will agree that editable results are a first step toward their more ambitious goal, and this enables them to cooperate with those who hold that mechanical translation need not and should not aim at anything more than the production of editable texts.

I have attempted to show in this paper that a base text in Interlingua is convertible by mechanical means into an editable translation in a target language belonging to the group of languages which are summarized in Interlingua. This does not imply that the same cannot be true for languages outside that group. It merely implies that such a more comprehensive assertion requires additional experimentation by competent investigators.

In any event, there is a group of languages (possibly quite extensive) which form a circle the center of which is occupied as it were by Interlingua. This suggests the possibility of utilizing Interlingua in mechanical translation as an intermediate language. A first step may have to be a more precise determination of what languages could be profitably involved in such a system. The second step would be the mechanization of the translation of texts written in Interlingua with all the links in its surrounding circle as target languages. If as a third step the reverse process of translating into Interlingua were likewise mechanized, all the languages in a group of $\underline{n}$ languages could be interconnected by $\underline{2n}$ processes of mechanical translation instead of by $\underline{n^2-n}$ such processes. The linking of twenty languages in all directions would not require three hundred and eighty processes but only forty.

Acknowledgment.-- The author's interest in problems of MT was first aroused by Professor Léon Dostert of Georgetown University. It matured through contact with Drs. Victor H. Yngve and William N. Locke of MIT. It turned to actual research thanks to the encouragement and guidance received from Dr. Yehoshua Bar-Hillel, likewise of MIT. Indebtedness to others, too numerous to list, is herewith acknowledged in cordial gratitude.

Bibliography

Yehoshua Bar-Hillel 58
Can Translation be Mechanized?
Methodos VII, No. 25-26, pp.45-62 (1955).

Reprinted from American Scientist (see abstract 41, Vol.1, No.2) with the addition of a two-page appendix bringing the article up to date.

V.H.Y.

Andrew D. Booth 59
Use of a Computing Machine as a Mechanical Dictionary
Nature, Vol.176, No.4481, p.565 (Sept. 17, 1955).

In a mechanical dictionary with the words stored in the numerical order according to their codes, it is proposed to find the desired word by a series of subtractions looking first in the middle, then at the 3/4 or 1/4 mark etc., instead of looking at each word in sequence until the right one is found.

V.H.Y.

J.P. Cleave and B. Zacharov 60
Language Translation by Electronics
Wireless World, Vol.61, No.9, pp.433-35 (Sept. 1955).

This paper gives a summary of the basic procedures to be employed in the machine translation of languages. Apparently, it is intended to clarify linguistic problems for engineers and some engineering problems for linguists.

J. Applegate

Jacob Ornstein 61
Mechanical Translation
Science, Vol.122, No.3173, pp.745-748 (Oct. 21, 1955).

Brief summary of the history of mechanical translation, with a more detailed account of the Georgetown-IBM experiment.

B. Shefts

Victor H. Yngve 62
Sentence-for-Sentence Translation
Mechanical Translation, Vol.2, No.2, pp.29-37 (Nov. 1955).

This paper is written for people with an information theory background. It discusses a suggested method for solving some of the multiple meaning and word order problems. Utilization of the sentence structure, that is grammar and syntax, may be able to solve more than the narrow grammatical and syntactic problems in translation.

Author

Abraham Kaplan 63
An Experimental Study of Ambiguity and Context
Mechanical Translation, Vol.2, pp.39-46 (Nov. 1955).

See abstract 3, Vol.1, No.1.

V.P. Berkov and B. A. Ershov 64
Concerning Attempts at Machine Translation
Voprosy Jazykoznanija, Nov.-Dec. 1955, pp.145-148. (In Russian).

Discusses the IBM-Georgetown demonstration.

V.H.Y.

William N. Locke 65
Translation by Machine
Scientific American, Vol.194, No.1, pp.29-33 (Jan. 1956).

Brief review article stressing cost to the U.S. of failure to keep abreast of Russian scientific output, need for high-speed, low-cost translations for coverage of foreign technical literature. Early development and present state of the art.

Author

Scientific Papers and Scientific Language 66
Nature, Vol.177, No.4497, pp. 1-2 (Jan. 7, 1956).

Report on the work done on problems of human communication under the support of the Nuffield Foundation. One of these is Booth's mechanical translation project at the Computational Laboratory of Birkbeck College.

B. Sheft

Published at the Massachusetts Institute of Technology, Cambridge, Massachusetts, supported in part by a grant from the Rockefeller Foundation, edited by Wiliam N. Locke*, Head of the Department of Modern Languages and* Victor H. Yngve*, Department of Modern Languages and research Laboratory of Electronics.*
Subscription price $1.00 for Volume Two (1955). Back issues are available at 50 cents apiece. Checks and money orders to be made payable to W.A.Hokanson*, Bursar, M.I.T. and sent to Mechanical Translation, Room 14N-307, M.I.T., Cambridge 39, Mass. Subscription correspondence should be addressed to Mechanical Translation at the same address.*
Editorial correspondence may be sent either to V.H.Yngve*, Room 20B-101B or* W.N. Locke*, Room 14N-307, both at M.I.T., Cambridge 39, Mass.*